The Face in Face-to-Face Communication

Signals of Understanding and Non-Understanding

ISBN: 978-90-76203-99-7

Cover design by Paul Hömke and Peter Nijland. Illustration by nchlsft (shutterstock).

Lenticular print by Lenticulair Benelux.

Printed and bound by Ipskamp Printing.

The Face in Face-to-Face Communication

Signals of Understanding and Non-Understanding

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus
prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 25 januari 2019
om 10.30 uur precies

door

Paul Hömke

geboren op 27 mei 1986
te Wetzlar, Duitsland

Promotor

Prof. dr. Stephen C. Levinson

Copromotor

Dr. Judith Holler

Manuscriptcommissie

Prof. dr. Asli Özyürek

Prof. dr. Herbert H. Clark (Stanford University, Verenigde Staten)

Prof. dr. Adrian Bangerter (Université de Neuchâtel, Zwitserland)

Dr. Anna Kuhlen (Humboldt-Universität zu Berlin, Duitsland)

Dr. Wyke J.P. Stommel

Het onderzoek in dit proefschrift werd mogelijk gemaakt door de Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Duitsland.

To Maria and Max, my little family

Table of Contents

Chapter 1. General introduction

> *As speakers, we are never indifferent to what the listener is doing. If he drums his fingers, if he frequently shifts in his chair, or looks about the room, or nods his head in an unusual pattern, he may convey the impression that he is bored, improperly attentive or inattentive, or that he is preoccupied. Sometimes this may throw us off balance to the extent that our flow of talk is brought to a stammering halt. More often, perhaps, we leave the encounter with a feeling of discomfort, with a feeling that there was no "rapport". However, we are usually unable to say what it is about the listener's behavior that gave us this feeling. Evidently we may be influenced by quite subtle features of his behavior.*
> (Kendon, 1990, p. 91)

What is the role of the listener in everyday communication[1]? Imagine you are sitting in a café, observing two people having a conversation. Chances are you would be looking at the person who is speaking, not the person who is listening. When observing conversations, people usually do not look at the person listening but at the person speaking (Augusti, Melinder, & Gredebäck, 2010; Holler & Kendrick, 2015; Casillas & Frank, 2017), much like, when watching a game of football, people tend to focus on the person who has the ball rather than on his or her defender. As such, studying listener behavior in face-to-face communication requires looking at the opposite side of where one is tempted to look, it requires a sort of "anti-saccade" (Hallet, 1978), literally but also metaphorically. Literally, because it requires resisting to look at the speaker while coding listener behavior in video corpora of conversations. Metaphorically, because studying listener behavior requires looking at phenomena that not many language scientists have previously looked at. Listener behavior seems to easily escape the radar of language scientists (Clark, 1997; Bavelas & Gerwing, 2011) as it seems to easily escape the radar of anyone else.

[1] Note that the terms 'listener' and 'addressee' are used interchangeably in this thesis. 'Addressee' is mainly used in Chapter 2 and 4. Keep in mind that, although the term 'listener' emphasizes the auditory modality, the 'listener' in conversation does not only listen but also produce visual signals (e.g., nods), just as the 'speaker' does not only speak but also produce visual signals (e.g., hand gestures).

In the language sciences, the role of the listener has been viewed from two main perspectives. In what Clark and Krych (2004) called "unilateral" views on language—widely adopted within linguistics and psycholinguistics—speaking and listening are individual processes, a view on language that is derived from more general, information-theoretic models of communication (e.g., Shannon & Weaver, 1949). Speakers determine the course of their utterances on their own, and listeners try to understand those utterances on their own. In psycholinguistics, processing models of speaking, for example, have been focusing on the selection of messages and the formulation and articulation of expressions (e.g., Garrett, 1980; Kempen & Hoenkamp, 1987; Levelt, 1989; Ferreira, 2000). While these models take into account that speakers monitor and repair their own speech (e.g., Levelt, 1983), they have "no provision for monitoring addressees and using that information to change course on line" (Clark & Krych, 2004, p. 63). Processing models of listening, on the other hand, have been focusing on parsing and interpreting utterances (Marslen-Wilson, 1987; Tanenhaus & Trueswell, 1995; Frazier & Clifton, 1996), rendering the listener as a passive receiver.

In "bilateral" views on language—widely adopted by conversation analysts (Sacks, Schegloff, & Jefferson, 1974; Schegloff, Jefferson, & Sacks, 1977; Goodwin, 1981; Atkinson & Heritage, 1984) and by an increasing number of psychologists and other cognitive scientists (Krauss & Weinheimer, 1966; Kraut, Lewis, & Swezey, 1982; Clark, 1996; Bavelas, Coates, & Johnson, 2000; Kopp, Allwood, Ahlsen, Grammer, & Stocksmeier, 2008; Brennan, Galati, & Kuhlen, 2010; Dale, Fusaroli, Duran, & Richardson, 2013)—language is considered a type of joint activity, an activity in which two (or more) people coordinate their actions moment by moment in order to achieve a shared goal (e.g., Clark, 1996; Sebanz, Bekkering, & Knoblich, 2006; see also Pickering & Garrod, 2004). In this view, speaker and listener coordinate moment by moment to achieve and maintain mutual understanding. According to bilateral views on language, the listener is an active collaborator. As Sacks and colleagues (1974) argued, "it is misconceived to treat turns as units characterized by a division of labor in which the speaker determines the unit and its boundaries", "the turn as a unit is interactively determined" (p. 727; see also Ford & Thompson, 1996; Schegloff, 1996).

The listener as an active collaborator

Whatever the speaker is doing in conversation, it is usually coordinated with what the listener is doing, much like the behavior of a football player having the ball is coordinated with his or her defender. Observational as well as experimental evidence supports the bilateral account of language and this collaborative nature of speaking and listening is particularly salient in grounding (Clark & Krych, 2004). Conversational participants ground their communicative acts by working together to establish the mutual belief that they have understood each other well enough for current purposes (Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Grounding is based on monitoring for communicative success and failure and on signaling understanding and non-understanding (or providing 'positive evidence of understanding' and 'negative evidence of understanding'; Clark, 1996).

To signal understanding, listeners produce feedback responses, such as *mm-hm* (see also 'back-channel' responses, Yngve, 1970), signaling successful grounding, which in turn affect the speakers' speaking (Bavelas, Coates, & Johnson, 2000; Clark & Krych, 2004; Malisz et al., 2016; Schegloff, 1982; Stivers, 2008; Yngve, 1970). In conversation analysis, a fine-grained classification of different types of listener feedback has been established based on the specific functions they fulfill in specific communicative contexts in conversation (Gardner, 2001). Listeners typically provide feedback at the end of speakers' turn constructional units (TCUs; Sacks et al., 1974). These are units with recognizable possible completions at which listeners can appropriately start a turn or initiate repair. To pass up the opportunity to take a turn at the end of a TCU, small behavioral tokens are produced by the listener (e.g., *mhm*), treating the current turn as still in progress, conveying "I'm with you, please continue", allowing the prior speaker to produce an extended turn consisting of multiple TCUs (Schegloff, 1982). Behaviors that serve this function in this position have been termed "continuers" (Schegloff, 1982; see also "markers of horizontal transition", Bangerter & Clark, 2003), and they have been contrasted with "epistemic tokens" (e.g., *oh*; Heritage, 1984), "activity-shift tokens" (e.g., *all right* or *okay*; Beach, 1993), and "assessments" (e.g., *oh wow*; Goodwin, 1986; see also "specific responses," Bavelas et al., 2000).

To signal non-understanding or to initiate repair, a different set of behavioral tokens can be produced by the listener (e.g., *huh?*), treating the prior unit as troublesome, signaling a problem in hearing or understanding, conveying "I'm *not* with you, please *don't* continue", inviting some sort of repetition or clarification (Sacks et al., 1974; Clark, 1996).

Vocal and visual listener feedback

Listener signals of understanding (continuers) and non-understanding (repair initiators) have been described in the vocal modality (e.g., *mhm*, *uh-huh*; Gardner, 2001; Schegloff, 1982 or *huh?*, Schegloff et al. 1977; Dingemanse et al., 2013). Language, however, is first acquired and primarily used in face-to-face contexts (Clark, 1996; Fitch, 2010) and—unlike other animals—humans tend to gaze at each other when communicating without necessarily signaling aggressive intent or sexual interest (Rossano, 2013; Holler & Levinson, 2014). The human's special status in the animal kingdom regarding the amount of mutual gaze, and more generally, the phylogenetic and ontogenetic centrality of face-to-face interaction for language use raises the question what role listeners' facial behavior plays in managing mutual understanding in conversation.

While the language sciences have made substantial progress in the study of hand gestures (especially of the speaker, e.g., McNeill, 2000; Kita & Özyürek, 2003; Kendon, 2004; Kelly, Özyürek, & Maris, 2010), the face has received relatively little attention despite its omnipresence in and intuitive relevance for everyday face-to-face communication. There is a large literature on *facial expressions* (Darwin, 1872; Ekman, 1993), which are often described as rather involuntary public manifestations of an individual's emotion, for example of fear upon seeing a spider. Facial expressions have been contrasted with more voluntary *facial gestures* (e.g., Bolinger, 1946). Facial gestures are facial actions that are used as communicative signals, that are shaped by the structure and content of a social interaction rather than by an individual's emotional response (Bavelas, Gerwing, & Healing, 2014a; see also "conversational facial signals", Ekman, 1979, and "facial displays", Kraut & Johnston, 1979). As Kendon (1975) suggested, "with regard to the face one asks, thus, not what feelings does the face express..., but what does the face do?" and "what functions for the interaction do... differentiable units of facial behavior have?", as (p. 300). Since 1975, more (but not much) research has been conducted on facial gestures. Speakers' facial gestures have been shown to serve depictions (Clark, 2016), for example to impersonate a character when telling a story (see also "reenactment", Sidnell, 2006; "facial portrayal", Bavelas, Gerwing, & Healing, 2014b; "multimodal quotation", Stec, Huiskes, & Redeker, 2015). Only a few studies have investigated listeners' visual bodily behavior, and only a small number of them have studied listeners' facial gestures. It has been shown, for example, that nodding (Heath, 1992; McClave, 2000; Stivers, 2008; Whitehead, 2011; Malisz, Włodarczak, Buschmeier, Skubisz, Kopp, & Wagner, 2016) but also facial behavior like smiling (Brunner, 1979) can signal understanding, and that eyebrow movements can signal non-understanding (e.g., Ekman, 1979; Manrique, 2015; Floyd et al., 2016)—if the visual bodily behavior in question is produced in the relevant sequential position (e.g., Mondada, 2011).

The eye region of the listener's face: Why investigate blinking and brow movements?

Now that several studies have established that listeners collaborate actively, not only vocally and by nodding, but also through facial behavior, a next step is to investigate more detailed aspects of listeners' facial behaviors and to explore their potential feedback function in face-to-face communication. What is the role of the listener's facial behavior in managing mutual understanding in face-to-face communication? What facial behaviors do listeners produce? And are speakers sensitive to these behaviors while speaking, i.e., do they influence speaker's linguistic behavior in face-to-face communication?

This thesis focuses on the eye region of the listener's face, since this region has been suggested to play an important role in mental-state signaling (Baron-Cohen, 2001; Lee et al., 2014), which is a crucial process for establishing and maintaining mutual understanding. While the first two empirical chapters investigate eye blinking as a potential signal of understanding, the last two empirical chapters focus on eyebrow movements and their role in signaling non-understanding.

Why is blinking interesting as a potential signal of understanding? Infants hardly blink (Ponder & Kennedy, 1927), but blink rate increases over time until adulthood (Zametkin, Stevens, & Pittman, 1979). Adults blink more frequently than physiologically required for ocular lubrication (Doane, 1980). We blink approximately 13,500 times every day—thus making it the most frequent facial action—with blinks being among the fastest movements the human body can generate (Peshori, Schicatano, Gopalaswamy, Sahay, & Evinger, 2001). In addition to reflex protective and physiological eye-wetting functions (Doane, 1980), blinking has been shown to index cognitive load. Under low cognitive load, people blink more than under high cognitive load (Ponder & Kennedy, 1927; Siegle, Ichikawa, & Steinhauer, 2008). A neuroimaging study has corroborated these behavioral findings showing that blinking deactivates the dorsal attention network while activating the default-mode network, suggesting an active involvement of blinking in attentional disengagement (Nakano, Kato, Morito, Itoi, & Kitazawa, 2013). Blinking has also been linked to social functions. Comparing different species of non-human primates has revealed that blink rate correlates positively with social group size (Tada, Omori, Hirokawa, & Ohira, 2013), which is a measure of social complexity associated with neocortex size (evidence used to support the "social brain hypothesis" Dunbar, 1992). Further, comparing different activity types in humans—that is, looking at a still target, reading, and conversing—the highest blink rate was found in conversation (Doughty, 2001). This suggests that in conversation blinking may fulfill functions beyond the physiological and cognitive. This thesis investigates whether listener blinking can serve communicative functions, specifically in signaling successful grounding in face-to-face communication (Chapters 2 and 3).

Why are eyebrow movements interesting as potential signals of non-understanding in spoken face-to-face conversation? Eyebrow movements are some of the most prevalent facial movements in conversation. In the emotion domain, eyebrow raises have been associated with positive emotions like surprise, and eyebrow furrows with negative emotions like anger (Ekman, 1993). In terms of non-emotional signaling, eyebrow movements have been thought to occur in requests for information from a conversational partner (Darwin, 1872; Eibl-Eibesfeldt, 1972; Ekman & Friesen, 1975; Ekman, 1979). Indeed, eyebrow position is a grammaticalized facial question marker in many sign languages (Baker-Shenk, 1983; Coerts, 1992; Zeshan, 2004; Dachkovsky & Sandler, 2009). Specifically, eyebrow movements have been shown to fulfill an important conventionalized function in signaling non-understanding in signed languages (Manrique, 2016). While problems in understanding can be initiated and resolved with spoken language in the absence of the visual channel (think of speaking on the phone), in spoken face-to-face conversation eyebrow raises and furrows have also been observed in signals of non-understanding (Enfield et al., 2013), but they might be epiphenomenal, that is, mere correlates or "ornaments" of verbal signals of non-understanding without a signaling function. This thesis investigates whether brow movements serve communicative functions in their own right, specifically in signaling non-understanding in spoken face-to-face communication (Chapters 4 and 5).

My general hypothesis is that listeners shape the progressivity, or forward movement, of the speaker's ongoing turn through facial feedback signals. On the one hand, some facial listener behaviors are hypothesized to signal understanding, "I've received enough information for current purposes, please go on" (e.g., long blinks), increasing the progressivity of the ongoing turn. On the other hand, other facial listener behaviors are hypothesized to signal "I've *not* received enough information for current purposes, please clarify" (e.g., eyebrow furrows), disrupting the progressivity of the ongoing turn by requesting some sort of repetition or additional clarification.

Methodology

To address the research questions of this thesis, I combined qualitative and quantitative corpus-based methods (Chapters 2 and 4) with experimental methods (Chapters 3 and 5). These methodological choices were based on several considerations.

As Levinson (2006, p. 39) pointed out:

> *Human interaction lies in an interdisciplinary no-man's land: it belongs equally to anthropology, sociology, biology, psychology, and ethology but is owned by none of them. Observations, generalizations and theory have therefore been pulled in different directions, and nothing close to a synthesis has emerged.*

Since 2006, an increasing number of interdisciplinary studies on human interaction have been conducted and they have contributed to an integration of observations, generalizations and theory of human interaction across disciplines (e.g., De Ruiter, Mitterer, & Enfield, 2006; Stivers et al., 2009; Dingemanse, Torreira, & Enfield, 2013; Keitel, Prinz, Friederici, Von Hofsten, & Daum, 2013; Holler, Kendrick, Casillas, & Levinson, 2016; Levinson, 2016; Holler, Kendrick, & Levinson, 2017). Nevertheless, as Levinson (2017) noted more recently, "the study of human communicative interaction is still in its infancy (...) and has only recently acquired the extensive public databases and measurements typical of cumulative science" (p.1). In an effort to further contribute to the advancement of an interdisciplinary science of human communicative interaction, I have selected different methodological tools from different disciplines—primarily from conversation analysis and experimental psychology—designed to balance ecological validity and experimental control and to look for converging evidence (de Ruiter & Albert, 2017).

Corpora of face-to-face conversation. As a starting point, we decided to use a corpus-based approach combining qualitative (conversation analytic) and quantitative methods because it allowed me to explore what listeners actually do with their faces in spontaneous face-to-face conversation and to generate functional hypotheses. We started with a corpus-based approach to get a detailed impression of listeners' facial behavior and to minimize the risk of potentially modeling phenomena experimentally at a later stage that are interactionally non-existent, extremely rare, or otherwise irrelevant in the sequential contexts of interest.

Two corpora of spontaneous, dyadic Dutch face-to-face conversations were used: my own purpose-built corpus of Dutch Face-to-Face (DF2F; see Chapter 2 and 4) conversation and the IFA Dialog Video Corpus (IFADV; van Son, Wesseling, Sanders, & Heuvel, 2008; see Chapter 4). Both corpora were specifically designed to allow for detailed analyses of communicative facial behavior.

The DF2F corpus consists of 10 dyads, all native Dutch speakers (18–68 years) who knew each other well prior to the recording. Four of the dyads consisted of a female and a male participant, four were all female, and two were all male. The dyads were engaged in spontaneous Dutch face-to-face conversations for 1 hour each and the recordings took place at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. The recordings took place in a soundproof room and participants were positioned approximately 1 m from each other at a 45-degree angle. Three HD video cameras (JVC GY-HM100) were used to record frontal views of each participant (see Supplementary Material 2) and a scene view. Audio was recorded using lightweight head-mounted microphones (DPA-d:fine-88). Each recording session consisted of three 20-minute phases. To achieve the highest audio quality for this study, only the 20-minute phase in which participants wore head-mounted microphones was used.

The IFADV Corpus consists of 23 dyads, all native Dutch speakers (12-72 years), who also knew each other well prior to the recording. Nine of the dyads consisted of a female and a male participant, 11 were all female, and three were all male. Five of the participants participated in two dyads each. The dyads were engaged in spontaneous Dutch face-to-face conversations for 15 minutes each. Conversations were recorded in a soundproof room, and participants were seated at a table, facing each other, positioned approximately 1 m from each other. Two video cameras (JVC TK-C1480B, 720x576, 25 fps) were used to record frontal views of each participant and audio was recorded using head-mounted microphones (Samson QV).

Since corpus data is correlational in nature, there is always the possibility that uncontrolled variables may play a role as well. Thus, as a next step, I used experimental methods (see Chapters 3 and 5) to test our observationally generated causal hypotheses. Although natural human language is multimodal and social-interactive in nature, traditional experimental approaches have primarily focused on verbal language and on utterances produced outside of a social-interactive context, allowing high experimental control but sacrificing ecological validity. As Levinson (2016) asked, “How can psycholinguistics investigate language in its native dialogic habitat?” (p. 12).

Virtual reality experimentation. This thesis presents a novel, virtual-reality-based experimental paradigm that embraces the multimodal and social-interactive nature of language while maintaining high experimental control. We decided to make use of humans’ “charitable over-attribution of interactional intelligence to anything that moves or squawks” (Levinson, 2017, p. 10). We decided to use virtual-reality technology since previous research has demonstrated that human agents can viably be replaced with virtual agents in research on social interaction as participants unconsciously attribute human-like characteristics to them (Nass & Moon, 2000; Casasanto, Jasmin, & Casasanto, 2010; Heyselaar, 2015). Questions regarding the causal role of subtle facial cues in interactive face-to-face communication have previously been impossible to address with a high degree of experimental control. The virtual-reality based experimental paradigm we developed, however, enabled me to selectively manipulate subtle facial behavior in a virtual listener, reducing problems regarding experimental control that would be inevitable when relying on a human listener (Kuhlen & Brennan, 2013).

My goal was to develop an experimental paradigm that is not only social-interactive on paper (see de Ruiter & Albert, 2017, for discussion) but that approximates as closely as possible the experience of being engaged in face-to-face conversation. One way I tried to achieve this was to emulate the relevant conversational sequence type as closely as possible (in this case, other-initiated tellings). Another was to have a human confederate control the timing of the listener feedback responses, as opposed to a computer program. A third way was to use pre-recorded human speech for the avatars' voices that was scripted but recorded in a social-interactive setting, which made them sound spontaneously produced rather then read from paper.

The only task for the participants in my experiments was to have a conversation with three different avatars and to respond to open questions (e.g., *How was your weekend, what did you do?*). While participants were answering, the avatar produced different types of visual feedback, which was triggered secretly by a confederate, who could see and hear the participant (via a video-camera link) but who was blind to the conditions and hypotheses and instructed to press a button whenever it felt appropriate to provide listener feedback. Crucially, the confederate's button presses triggered different types of visual feedback responses in the avatar, which were varied across conditions, allowing us to test my observationally generated hypotheses regarding how different types of visual listener feedback may influence the linguistic behavior of speakers in face-to-face communication.

Overview of thesis

The following four chapters present the four empirical studies conducted for this thesis, as outlined below. Note that since some of the empirical chapters are based on published articles or submitted manuscripts, some portions overlap with this general introduction chapter.

Chapter 2 investigates whether eye blinking might function as a type of listener feedback. To explore this possibility, we built a corpus of spontaneous, informal, dyadic Dutch face-to-face conversations, identified short and long listener blinks during extended turns, and measured their occurrence relative to the end of speaking units, that is, relative to the end of turn constructional units (TCUs), the location where feedback typically occurs. Listener blinks were indeed timed to the end of TCUs. Also, long blinks were more likely than short blinks to occur during mutual gaze, with nods or continuers, and their occurrence was restricted to sequential contexts in which signaling understanding was particularly relevant, suggesting a special capacity of long blinks to signal "I've received enough information for current purposes”. In the same way in which brow furrowing (as if trying to see more clearly) seems to signal a lack of understanding, closing the eyes by blinking may signal “no need to ‘see’ anymore” because sufficient understanding has been reached.

Chapter 3 investigates whether speakers are sensitive to listener blink behavior as a communicative signal, that is, whether the speaker’s linguistic behavior is influenced by listener’s blink behavior in face-to-face communication. Chapter 3 builds on the correlational findings from Chapter 2 and experimentally tests the observationally generated hypothesis that listener blink behavior is taken into account by speakers and that it serves a communicative feedback function signaling “I’ve received enough information for current purposes”. To test this hypothesis, we used virtual reality to develop a novel experimental paradigm enabling us to selectively manipulate blinking in a virtual listener, crucially distinguishing between short and long blinks. It was found that speakers unconsciously took into account small differences in listener blink duration, producing shorter answers in the context of long listener blinks, apparently perceiving these as signaling “I’ve received enough information for current purposes”. These findings demonstrate for the first time that, in addition to physiological, perceptual and possible cognitive functions, listener blinking may serve as a feedback signal in interactive face-to-face communication. More generally, these findings may shed new light on the visual origins of mental-state signaling, which is a crucial ingredient for achieving mutual understanding in everyday social interaction.

Chapters 2 and 3 investigate listener’s eye blinking as a potential signal of understanding. The next chapters, Chapters 4 and 5, also focus on the eye region of the listener’s face, investigating eyebrow movements as potential signals of non-understanding.

Chapter 4 investigates the role of eyebrow movements in signaling communicative problems. Are eyebrow raises and furrows functionally involved in signaling problems of hearing or understanding or might they be epiphenomenal, mere correlates of verbal signals of problems in hearing or understanding? To address these questions, I collected data from two corpora of face-to-face Dutch conversations, coded the co-occurrence of eyebrow movements with different types of verbal signals of problems in hearing or understanding (or repair initiations), the temporal relationship between the visual and verbal component in these multimodal signals, the type of solutions provided in response, and eyebrow movements alone that were treated as signals of problems in hearing or understanding. It was found that, while eyebrow raises and furrows co-occurred with all basic types of verbal repair initiations, verbal signals co-occurring with a brow furrow were more likely to be responded to with clarifications compared to verbal signals co-occurring with a brow raise or no brow movement at all. Second, when speakers were forewarned visually through a brow movement by their recipient that a verbal repair initiation would come up, communicative problems were solved faster than if they were not forewarned through a brow movement, suggesting that brow movements may enhance communicative efficiency. Finally, while brow movements were not necessary for initiating repair, brow furrows alone appeared to be sufficient, suggesting a unique capacity of brow furrows to signal "I've not received enough information for current purposes"—without relying on words. These findings suggest that brow movements are communicative signals in their own right, and that brow raises and furrows may fulfill partially different functions. More generally, they suggest that brow raises and furrows go beyond expressing emotions, and that they are frequently used for signaling informational needs in everyday communication.

Chapter 5 investigates whether speakers are sensitive to listener brow furrows as a communicative signal. It builds on the correlational findings from Chapter 4 and experimentally tests the observationally generated hypothesis that listener's eyebrow furrowing is taken into account by speakers and that it serves a communicative feedback function signaling "I've *not* received enough information for current purposes". To test this hypothesis, we used virtual reality to selectively manipulate eyebrow furrowing in a virtual listener (see Chapter 3 for similar methods). It was found that speakers produced longer answers when talking to a brow-furrowing listener than when talking to a listener that nodded throughout, thus supporting the hypothesis that listener eyebrow furrowing can indeed signal insufficient understanding. The differences in answer length could neither be alternatively explained by differences in hesitations, nor by differences in speakers' perception of how human or 'natural' the virtual listeners appeared as conversational partners in the different conditions. Taken together, these results suggest that speakers may incorporate listeners' brow behavior into their recipient design, treating listener brow furrows as signaling "I've *not* received sufficient information for current purposes" by providing additional information. Thus, in addition to visual, emotional, and possible cognitive functions, brow furrows may serve as cooperative signals of insufficient understanding. Like the findings on eye blinking, the findings on eyebrow furrowing highlight the importance of the eye region for mental-state signaling, a crucial ingredient for achieving intersubjectivity in everyday communication.

Chapter 6 summarizes the main findings, discusses theoretical implications, and highlights potential avenues for future research.

Signaling understanding facially

Chapter 2. Eye blinking as addressee feedback in face-to-face conversation[2]

Abstract

Does blinking function as a type of feedback in conversation? To address this question, a corpus of Dutch conversations was built, short and long addressee blinks were identified during extended turns, and their occurrence was measured relative to the end of turn constructional units (TCUs), the location where feedback typically occurs. Addressee blinks were indeed timed to the end of TCUs. Also, long blinks were more likely than short blinks to occur during mutual gaze, with nods or continuers, and their occurrence was restricted to sequential contexts in which signaling understanding was particularly relevant, suggesting a special signaling capacity of long blinks.

[2] Hömke, P., Holler, J., & Levinson, S. C. (2017). Eye blinking as addressee feedback in face-to-face conversation. *Research on Language and Social Interaction*, *50*(1), 54-70.

Introduction

Language is primarily used in face-to-face conversation (Clark, 1996). The role of the addressee in conversation has been viewed from two main perspectives in the language sciences. In what Clark and Krych (2004) called “unilateral” views on conversation—widely adopted within linguistics and psycholinguistics—speaking and listening are individual processes. Speakers determine the course of their utterances on their own, and addressees try to understand those utterances on their own. In this view, the addressee is a passive receiver. In “bilateral” views on conversation—widely adopted by conversation analysts and some psychologists—speaking and listening is considered a joint activity in which speaker and addressee coordinate moment by moment to maintain mutual understanding (Brennan, Galati, & Kuhlen, 2010; Clark, 1996; Goodwin, 1981; Sacks, Schegloff, & Jefferson, 1974). According to this view, the addressee is an active collaborator. Observational as well as experimental evidence supports the bilateral account of conversation: While speakers are speaking, addressees provide vocal feedback like *mm-hm* and visual feedback like nods and smiles, which in turn affect the speakers’ speaking (Bavelas, Coates, & Johnson, 2000; Brunner, 1979; Clark & Krych, 2004; Malisz et al., 2016; Schegloff, 1982; Stivers, 2008; Yngve, 1970).

In conversation analysis, a fine-grained classification of different types of addressee feedback has been established based on the specific functions they fulfill in specific sequential positions in conversation (Gardner, 2001). Addressees typically provide feedback at the end of speakers’ turn constructional units (TCUs; Sacks et al., 1974). These are units with recognizable possible completions at which next speakers can appropriately start a turn or initiate repair. To pass up the opportunity to take a turn or initiate repair at the end of a TCU, small behavioral tokens (e.g., mhm) are produced by the addressee that treat the turn as still in progress, conveying “I’m with you, please continue,” allowing the prior speaker to produce an extended turn consisting of multiple TCUs (Schegloff, 1982). Behaviors that serve this function in this position have been termed “continuers” (Schegloff, 1982), and they have been contrasted with “epistemic tokens” (e.g., oh; Heritage, 1984), “activity-shift tokens” (e.g., all right or okay; Beach, 1993), and “assessments” (e.g., oh wow; Goodwin, 1986; see also “specific responses,” Bavelas et al., 2000). While continuers have been primarily described in the vocal modality (e.g., *mhm*, *uh-huh*; Gardner, 2001; Schegloff, 1982), it has been shown that visual conduct—nods in particular (Heath, 1992; McClave, 2000; see also Stivers, 2008; Whitehead, 2011)—can also serve a continuer function if produced in the same sequential position (e.g., Mondada, 2011).

The goal of this study was to investigate blinking as one potential additional type of visual addressee feedback and more specifically, as visual conduct potentially serving a continuer function. Humans hardly blink at birth (Ponder & Kennedy, 1927), but blink rate increases until adulthood (Zametkin, Stevens, & Pittman, 1979). Adults blink more often than physiologically necessary for wetting the eyes (Doane, 1980), showing a blink rate of approximately 15 to 20 blinks per minute, with a mean blink duration of 300–400 ms (e.g., VanderWerf, Brassinga, Reits, Aramideh, & deVisser, 2003).

In terms of function, blink rate has been shown to index cognitive load. People blink less under high cognitive load and more under low cognitive load (e.g., Nakano, Yamamoto, Kitajo, Takahashi, & Kitazawa, 2009; Siegle, Ichikawa, & Steinhauer, 2008). Supporting these behavioral findings, a neuroimaging study has revealed that blinking activates the default-mode network while deactivating the attention network, suggesting an active involvement of blinking in attentional disengagement (Nakano, Kato, Morito, Itoi, & Kitazawa, 2013). At the same time, blinking has been linked to social-communicative functions. Looking at different activity types in humans—staring at a target, reading, having a conversation—the highest blink rate was found in conversation (Doughty, 2001). In nonhuman primates, blink rate is correlated with group size (Tada et al., 2013), a measure of social complexity that has been linked to neocortex size (evidence used to support the "social brain hypothesis," Dunbar, 1992). These findings suggest that in addition to peripheral physiological and central cognitive functions, blinking may serve a social-communicative function (see also Mandel, Helokunnas, Pihko, & Hari, 2014; Nakano & Kitazawa, 2010; Tada et al., 2013).

To my knowledge, there are only two studies specifically investigating addressee blinking in conversation. Sultan (2004) demonstrated that in American Sign Language addressees use blinks to signal understanding: "I'm with you. I've got it. You can continue." (Sultan, 2004, p. 50). She suggests that addressee blinking may have developed a feedback function in signed languages because of the need to control blinking to minimize visual information loss. However, addressee blinking as a signal of understanding has also been described in Yélî Dnye, a spoken language of Papua New Guinea (Levinson & Brown, 2004).

In the present study, it was hypothesized that addressee blinking may serve a similar function in spoken Dutch, which also relies heavily on the visual channel—like most spoken languages do, at least in face-to-face contexts[3]. If addressee blinking does indeed fulfill a feedback function in spoken conversation, one should expect addressee blinks to be timed to speakers' talk at similar points in time as other addressee responses, namely, at the ends of syntactically, prosodically, and pragmatically complete units (Ford, Fox, & Thompson, 1996; Gardner, 2001; Schegloff, 1982; Selting, 2000; Yngve, 1970). If addressee blinking does not serve a feedback function in spoken conversation, one may expect addressee blinks to be randomly distributed across turns, as their occurrence should not be influenced by conversational context.

To address this question, a corpus of dyadic Dutch face-to-face conversations was built. Then, the timing of addressee blinks was quantified by measuring the temporal distance of each blink onset to the closest TCU end (focusing on syntactically, prosodically, and pragmatically complete units within extended turns). In addition, the multimodal compositionality of addressee blinks was examined as well as the placement of long addressee blinks in conversational context.

Methods

Participants and corpus

The corpus consists of 10 dyads engaged in spontaneous Dutch face-to-face conversations for 1 hour each. All participants were native Dutch speakers (18–68 years; mean age = 30.7), they knew each other prior to the recording, and each participant participated only in one dyad. Four of the dyads were all female, four consisted of a female and a male participant, and two were all male. The recordings took place at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands.

Setup and equipment

The conversations were recorded in a sound-proof room, participants were positioned approximately 1 m from each other at a 45-degree angle, and each participant wore a

[3] But see speakers of Mayan Tzeltal for an exception (e.g., Rossano, Brown, & Levinson, 2009).

head-mounted microphone (DPA-d:fine-88). Three HD video cameras (JVC GY-HM100) were used to record frontal views of each participant and a scene view (see Supplementary Material 1). An audio recorder (Roland R-44) recorded the two audio tracks in synchrony. Each recording session resulted in three videos and two audio files, which were then synchronized and exported in Adobe Premier Pro CS6 (MP4, 24 fps).

Procedure

Each recording session consisted of three 20-minute phases. To achieve the highest audio quality for this study, only the 20-minute phase in which participants wore head-mounted microphones was used. The whole session lasted about 90 minutes, and each participant was paid 16 euros. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen, and informed consent was obtained before and after filming.

Analysis

Turns analyzed. We selected tellings that occurred naturally in the conversations, and of these we selected only those that were elicited by the other speaker ("second-position tellings"; Mandelbaum, 2013). To address the question of whether addressee blinking may have a feedback function, we treated the whole telling, over several syntactically, prosodically, and pragmatically complete TCUs, as a single turn and excluded tellings with "recipient disruptions" (Mandelbaum, 2013), that is, nonminimal recipient responses (larger than three syllables), and other-initiations of repair or news receipts (e.g., Oh did you?) since they make a speaker response relevant before the telling continues[4].

Turns and points of possible completion. Within all selected 46 turns, 456 points of possible completion were annotated, marking boundaries of *final* (230) and *nonfinal* (226) TCUs. At any point at which a turn was hearable to the coders as

[4] Prior research suggests that blinking behaves differently in speakers as opposed to addressees (Cummins, 2012). Since I was interested in addressee behavior, I chose extended turns as a starting point because—compared to more turn-by-turn interaction—they provide a relatively clear distinction between speaker and addressee roles. The focus on extended turns in second position was based on plans for a follow-up experiment in which participant tellings would be initiated by an experimenter and for which this corpus study should serve as an observational basis.

possibly complete in its context, focusing on convergence of syntactic, prosodic, and pragmatic completion, a *final* point of possible completion (Selting, 2000) was annotated, without this necessarily constituting the actual turn end. Final points of possible completion thus correspond to what Ford et al. (1996) called “Complex Transition Relevance Places,” places at which speaker transition may occur.

Nonfinal points of possible completion were annotated if an utterance in progress was hearable to the coders as locally syntactically complete (e.g., a complete clause, whether syntactically independent or dependent) and locally prosodically complete (i.e., a complete intonational phrase) but further talk was made projectable syntactically (e.g., as in the case of an if clause), prosodically (e.g., rising intonation), or pragmatically (e.g., an incomplete answer to a question). Although at these nonfinal points of possible completion further talk is made projectable, these are typical points for addressee responses (“local pragmatic completion,” Ford et al., 1996; Lerner, 1996). Whether addressee feedback is timed to nonfinal or final points of possible completion may have functional implications. For example, while feedback timed to nonfinal points may pass up the opportunity to initiate repair, feedback timed to final points may additionally pass up the opportunity to take a full turn (see also Goodwin, 1986).

The main coder (PH) and a second coder (EV) who was blind to the visual context in which the turns were produced identified the location of final and nonfinal points of possible completion by listening to the audio. Thirty-one extended turns were coded for training, and 11 extended turns (i.e., 24%) were coded to measure reliability. Initial discrepancies in coding regarded primarily prosodic completion points, and consensus was achieved through thorough discussion of cases where the two coders disagreed. Evaluation of their coding revealed high intercoder agreement on the identification of nonfinal and final TCU ends (84%).

Addressee blinks. Participants’ blinks were detected automatically using motion tracking software (“IntraFace”; Xiong & De la Torre, 2013; see Figure 1 and Supplementary Material 2 for a video example).

Figure 1. Still image from facial motion tracking-based blink detection. The green little dots on the green lines represent the pixels being tracked over time. A blink was automatically detected when the distance between dots on the upper lid and dots on the lower lid was reduced beyond a predefined threshold.

Within second-position tellings, all detected addressee blinks were manually corrected in terms of false positives, false negatives, and blink duration. Blink duration annotations included the first frame in which a downward movement of the eyelids was observable and the last frame in which an upward movement of the eyelids was observable[5]. Voluntary blinks have been shown to have longer durations (Kaneko & Sakamoto, 1999) and according to Levinson and Brown (2004), and a neuroimaging study (Mandel et al., 2014), longer blinks have a special communicative salience. We therefore decided to categorize blinks into short and long blinks. Previous research categorized blinks based on duration using a threshold of 250 ms (Levinson & Brown, 2004), 240 ms (and 400 ms; Cummins, 2012), and 420 ms (Hermann, 2010). We used a similar threshold as Hermann (2010), who, as we did in the present study, used the first observable downward movement of the eyelids to determine the blink onset and the last observable upward movement to determine the blink offset. We used a threshold of 410 ms, which separated the longest 25% from the rest (splitting them at the upper quartile[6]). This resulted in 350 short blinks (<410 ms) and 61 long blinks (≥410 ms).

[5] Our criterion mirrors Hermann (2010) but contrasts with Cummins (2012), who did not annotate a blink onset before the visible part of the cornea was not "substantially occluded" (p. 8), resulting in shorter blink durations overall.

[6] I split them at the upper quartile rather than the median because using the median as a threshold would have included too many blinks that were impressionistically short blinks in the category of "long blinks." What seems to differentiate impressionistically short from

Multimodal compositionality. We assessed for each addressee blink co-occurring addressee behaviors. We focused our analysis on the most salient addressee responses, namely, on nods (vertical head movement including at least one upward and one downward movement), vocal continuers (e.g., mm-hm), and combinations of these. Blinking was considered to be co-occurring if the blink overlapped with a nod or a vocal continuer or if it preceded or followed the nod or vocal continuer without perceived interruption, such that the behaviors together formed a multimodal Gestalt (Mondada, 2014). Blinking was not considered to be co-occurring if there was a temporal distance ≥ 250 ms between the blink and the nod or the vocal continuer. Secondly, blinking was considered with respect to the interactants' gaze direction. Blinking was considered as having occurred during a period of mutual gaze if mutual gaze existed at the onset of the blink, leading to a disruption of mutual gaze as the blinking was executed. All annotations were created in ELAN 4.8.1 (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006).

Results

The analysis is focused first on all addressee blinks and then examines potential differences in timing in different subsets of the data (short versus long blinks, blinks timed to nonfinal versus final TCUs). We then compare the multimodal compositionality of short and long blinks, followed by quantitative analyses with qualitative analyses of long addressee blinks in conversational context.

Addressee blinks timed relative to the end of TCUs

First, we looked at the frequency of addressee blinks. Addressees blinked approximately every other second on average, although there was substantial interindividual variability (mean blinks per minute = 30.53; *SD* = 20.43). Then we measured the temporal distance (in ms) between each blink onset and the closest end of a TCU (collapsing across short and long blinks and nonfinal and final TCU ends). If blink timing was random and TCU duration was constant, one would expect a uniform distribution of blink timings, that is, a flat horizontal line, indicating equal likelihood of blink occurrence at any point during a TCU. However, the analysis

---

impressionistically long blinks is not so much the duration of the opening phase, nor the duration of the closing phase, but rather the duration of the cornea being completely occluded.

revealed that the most typical timing of blinks was very close to the TCU end (estimated mode = 52 ms, median = 20 ms, mean = −20 ms; see Figure 2), especially when considering the average TCU length of 1754 ms. This suggests that addressees tend to coordinate the onset of their blinks with the end of a speaker's TCU (see Supplementary Materials 3a and 3b for video examples).

However, an alternative explanation needs to be ruled out. TCU duration was variable, with many relatively short and much fewer long TCUs. Thus, even if blinks were timed randomly with respect to the end of TCUs, the temporal distance between blink onset and TCU end would necessarily be smaller for shorter than for longer TCUs. This, too, could lead to a distribution that peaks around zero.

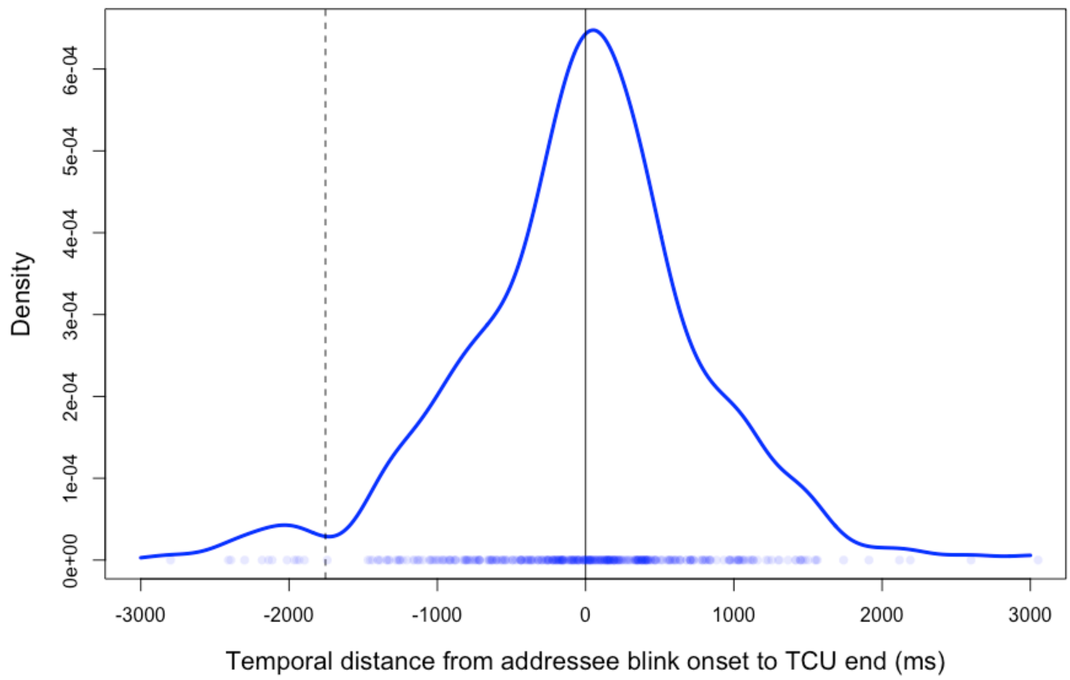

Figure 2. Addressees' blink onset (N = 411) relative to the closest TCU end. The vertical line at the zero point of the x-axis marks the TCU end. The distance between this line and the dashed vertical line marks the average TCU duration (1,754 ms). The peak of the distribution represents the estimate of the mode (52 ms).

To rule out this alternative explanation, we standardized the data by dividing the temporal distance between each blink onset and the closest TCU end by the duration of the respective TCU. This provided a measure indicating how closely each blink was timed to the TCU end relative to the duration of that TCU. A value of 0 means that the blink onset coincided with the TCU end, and a value of 1 means that the temporal distance of blink onset to TCU end was as large as the TCU itself, that is, the blink onset occurred at the beginning of the TCU. Visual inspection of the data reveals that taking into account the variability in TCU duration increased the spread of the distribution (see Figure 3) compared to the unstandardized data, which shows a relatively tighter distribution around zero (see Figure 2).

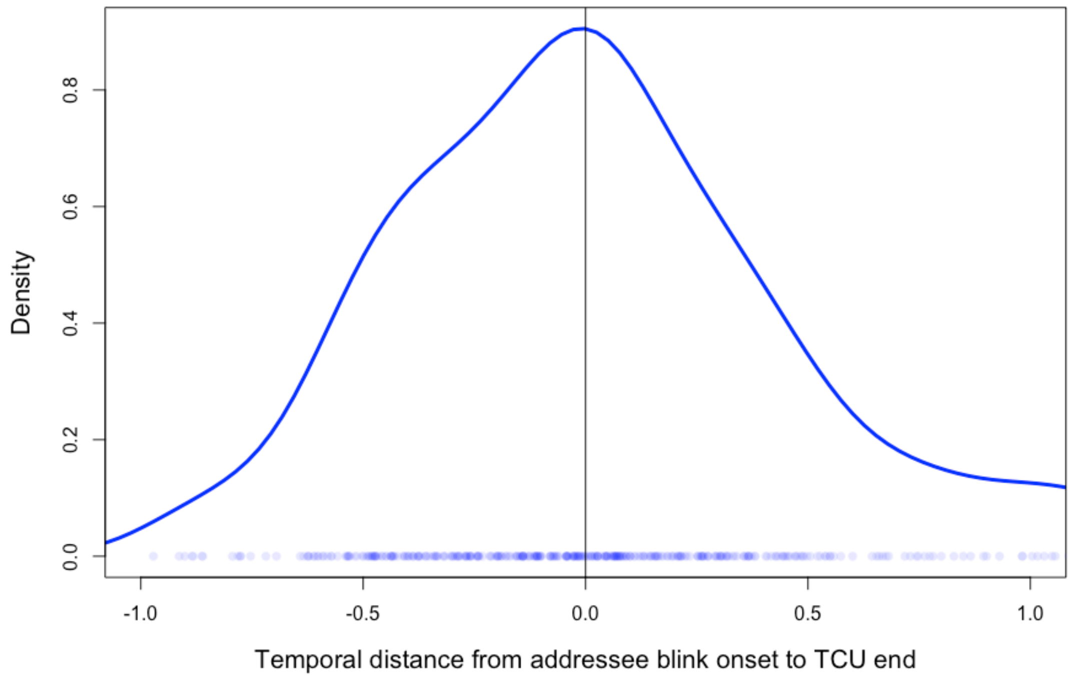

Figure 3. Timing of addressee blink onset relative to TCU ends based on standardized TCU duration.

However, measures of central tendency of the standardized data revealed that blinks were again most typically timed close to the end of TCUs (estimated mode = .00, median = .01, mean = .20). Taken together, this points to a clear tendency of addressees to coordinate their blinking with the end of TCUs.

Short versus long addressee blinks timed relative to nonfinal and final TCU ends. We used R (R Core Team, 2012) and lme4 (Bates, Maechler, & Bolker, 2012) to test in a mixed effects model whether standardized blink timing differed depending on blink duration (short versus long blinks) and type of unit end (nonfinal versus final). Outliers deviating more than two standard deviations from the mean (n = 11, 2.6 % of the data) and three participants contributing less than five data points (requirement for lme4) were excluded. As fixed effect, we entered blink duration (short, long) and finality of TCUs (nonfinal, final), plus the interaction term. Participants were modeled as nested inside conversations, and as random effects we entered intercepts for participants into the model. Main effects and interaction effects were calculated using the ANOVA function of the car-package (Fox & Weisbert, 2011). Although based on visual inspection of Figure 4, long blinks seem to have occurred later than short blinks relative to TCU ends, this difference is not statistically significant (main effect of blink duration, $\chi 2(1) = 2.45$, $p = .11$). Moreover, blinks were not timed significantly differently relative to nonfinal versus final TCU ends (main effect of type of unit end, $\chi 2(1) = .32$, $p = .56$, and neither were short and long blinks timed significantly differently at nonfinal versus final TCU ends (interaction of blink duration and type of unit end, $\chi 2(1) = .10$, $p = .74$).

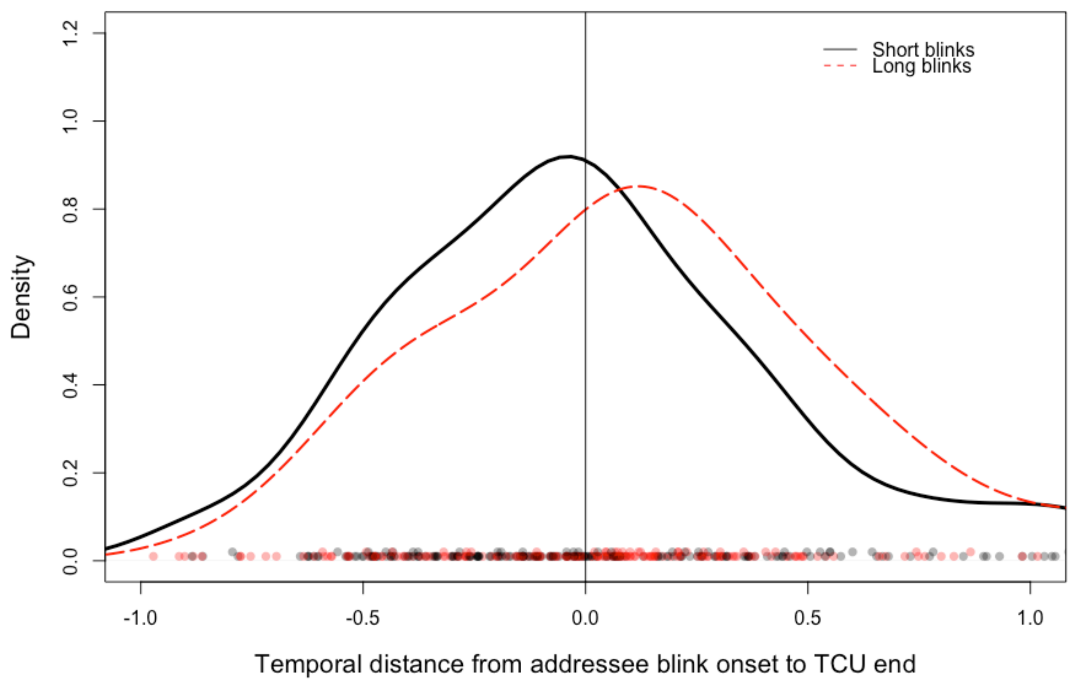

Figure 4. Addressee blink onset of *short* blinks versus *long* blinks relative to the ends of TCUs standardized in duration.

Multimodal compositionality of short vs. long addressee blinks

While short and long blinks were not timed differently relative to TCU ends, they might differ in the extent to which they co-occur with other addressee responses. We focused on the most salient co-occurring addressee responses in our corpus, namely, on nods, vocal continuers (e.g., *mm-hm*), and combinations of these (see Method section). As one can see in Figure 5, while most short blinks (73% [n = 256]) and approximately half of the long blinks (54% [n = 33]) occurred without other responses (i.e., neither with nods nor with vocal continuers), long blinks were approximately twice as likely to co-occur with nods (18% [n = 11]) or vocal continuers (18% [n = 11]) than short blinks (9% [n = 32] and 8% [n = 29] respectively), supporting impressionistic observations by Cummins (2012).

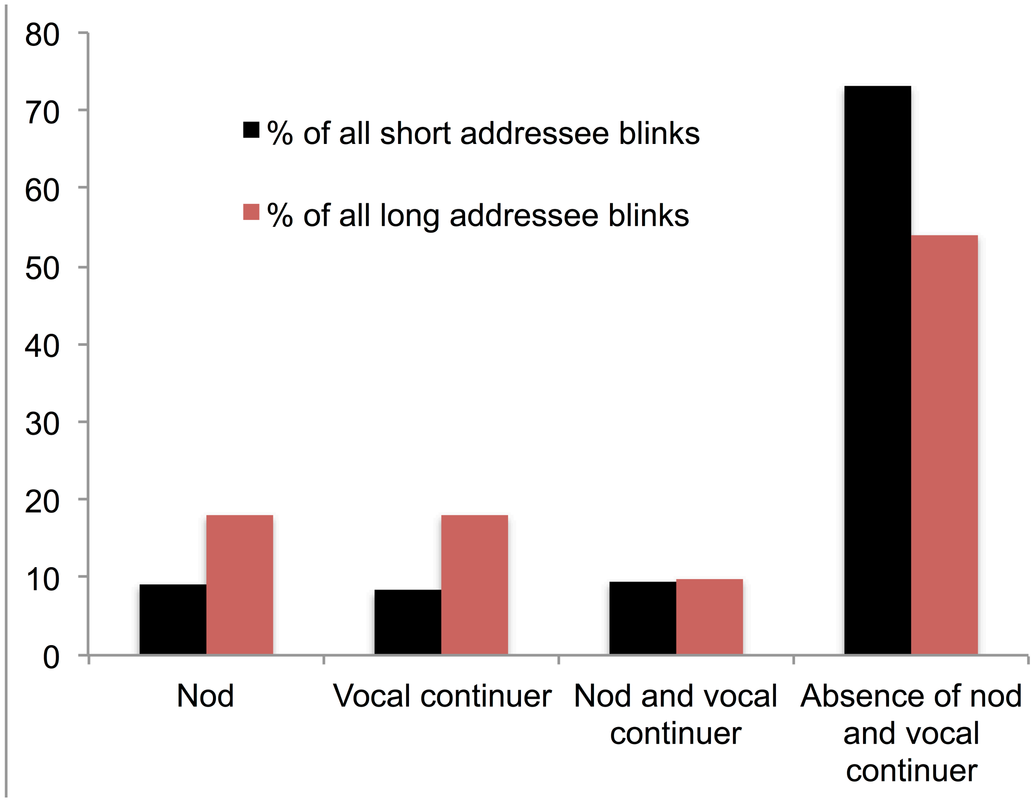

Figure 5. Co-occurrence of short blinks (in black; $n = 350$) versus long blinks (in red; $n = 61$) with nods, continuers, nods combined with continuers, and with no other addressee response.

Predicting blink duration (short, long) by co-occurrence of other addressee responses (nod, vocal continuer, nod plus vocal continuer, absence of nod and vocal continuer) in a mixed effects logistic regression analysis (including random intercepts for participants) confirmed this observation statistically, indicating that the co-occurrence of other addressee responses reliably distinguished between short and long blinks, $\chi2(3) = 11.6$, $p = .001$. Interestingly, long blinks were as likely to co-occur with vocal continuers and nods (10% [n = 6]) as short blinks (9% [n = 33]) were. One possibility is that with vocal continuers, the long blink compensated for the lack of a nod, and with nods, the long blink compensated for the lack of a vocal continuer. When nods and vocal continuers were combined into a "composite signal" (Clark, 1996), their joint "continuer effects" might have been sufficient, making the addition of a long blink more optional.

An additional analysis revealed that long blinks were also significantly more likely to be produced during the mutual "gaze window" (Bavelas, Coates, & Johnson, 2002), namely 72% of long blinks compared to 47% of short blinks[7]. Predicting blink duration (short, long) by mutual gaze (present, absent) in a mixed effects logistic regression analysis (including random intercepts for participants) confirmed this observation statistically, indicating that the presence of mutual gaze reliably distinguished between short and long blinks, $\chi 2(1) = 9.7$, $p < .001$.

Overall, we would like to conclude from these patterns that (a) addressee blinks cluster at both final and nonfinal TCU ends; (b) that long blinks have a distinctive distribution regarding the concurrent production of nods and continuers; and that (c) long blinks are more likely than short blinks to be produced during mutual gaze—together suggesting that long blinks have a signaling capacity, producing feedback at critical points in the telling.

The sequential placement of long addressee blinks

The goal of this section is to further address the question of whether long blinks can serve a communicative function by examining the sequential contexts in which they were produced. To achieve this goal, we will first present a quantitative analysis of long blinks in different sequential contexts. Following this, we will provide qualitative analyses suggesting that long blinks can serve a continuer function (especially in combination with nods or vocal continuers) by displaying recipiency, passing up the opportunity to take a full turn or to initiate repair (Robinson, 2014; Schegloff, 1982) while signaling successful grounding (Clark, 1996). Finally, we will discuss potential alternative interpretations.

While short blinks occurred at a wide range of sequential locations in conversation, the use of long blinks was restricted to specific sequential contexts, namely, in

[7] For this analysis, all 61 long blinks were compared with 61 randomly selected short blinks.

response to repair solutions in self-initiated, same-turn, self-repair (47.5%), in response to disfluent speech (e.g., *uh, uhm*; 19.7%), at early recognition points (16.4%), and referring expressions (8.2%)[8,9].

Comparing all 61 long blinks to 61 randomly selected short blinks revealed that short blinks were significantly less likely to occur in these same contexts: in response to repair solutions in self-initiated, same-turn, self-repair (6.5%); at early recognition points (8.2%); in response to disfluent speech (3.2%); and referring expressions (1.6%). Predicting blink duration (short, long) by sequential context (self-repair, disfluency, early recognition point, referring expression) in a mixed effects logistic regression analysis (including random intercepts for participants) confirmed this observation statistically, indicating that the sequential context reliably predicted blink duration, $\chi 2(4) = 45.9$, $p = .001$. The following examples focus on the two sequential contexts in which long blinks most frequently occurred, self-repairs and disfluencies.

The long addressee blink in response to repair solutions in self-initiated, same-turn, self-repair. About half of the long blinks occurred after speakers' self-initiated, same-turn, self-repairs (Schegloff, 2013; Schegloff, Jefferson, & Sacks, 1977). Addressee responses that can occur in this position are vocal continuers (Mazeland, 2007), nods (Healey, Lavelle, Howes, Battersby, & McCabe, 2013), or combinations of these. The following extracts demonstrate that long blinks also occur in this position, typically combined with nods or vocal continuers. In the transcripts, addressee blinks are bracketed between two small dots [·] and are synchronized with corresponding stretches of talk. Each large dot [●] reflects approximately 100 ms, so four [●●●●] or more large dots refer to a "long blink" (see the appendix for a full description of the transcription conventions used).

In Extract 1, A and B had talked about experiments to participate in at the

---

8 Note that by "response" and "in response to" I do not intend to claim that long blinks, vocal continuers, etc., are necessarily conditionally relevant in these environments (since the present study does not demonstrate this), but see Zama & Robinson (2016) for further discussion.

9 By "early recognition points" I refer to points at which it becomes highly predictable for the addressee what the speaker intends to say. They can occur within clauses where remaining words or syllables are highly predictable, but they can also occur at preliminary component completions within compound TCUs where the remaining units of the compound are highly predictable (e.g., in if-then constructions; Lerner, 1996). By "referring expressions" I refer to noun phrases identifying individual objects, events, or beings (e.g., proper names like John).

university when A asks about a particularly long one B had previously mentioned (line 1).

```
Extract 1 (ETC_16_75310)

01 A: Maar wat was die andere nou? (.) tweeënhalf uur nog wat
      but what was the other one now  two.and.a.half hour something
      But what was the other one now?  Two and a half hour something

02    (0.3)

03 B: +Ja en dan moest je-dan krijg je zo'n m+iddel toegediend?
      yes and then must you then get you such one agent administered
      Then you get some agent administered?
      +gaze averted                           +


04   T·MS· of zoiets,    [Ik weet niet pr]ecies,
      TMS or such something I know not
      precisely TMS or something, I don't know
      exactly,
   a  ·●●●·

05 A:                    [Ja],
                          yes
                          Yeah

06 B: En dan werden je hersenen worden dan gestimuleerd .h,
      and then were your brains   are  then  stimulated
      And then your brains were then stimulated .h,

07    (0.2)

08 a   [°mhm°  ]
      ·●●●       .


09 B: ·[Of bepa]·alde:: onderdelen van je hers·e#ne*·.hhh*hhh,
        or certain      parts      of your brains
        Or certain parts of your brain
→ a                                         ·    ·●●●●
                                                      *nod  *
 fig                                              #fig.6


10 B: En+-eh: dan gaan ze ook allemaa::l (0.4) testjes met +j·e doen=
      and     then go they also all sorts of tests  with you do
      An- uh: then they go and do all sorts of tests with you
        +gaze averted                                      +
   a                                                   ·●●●->
```

```
11 B: =ik· weet niet precies.
       I   know  not precisely
       I don't know exactly.
    a   ->·
```

Figure 6. Addressee's long blink and head nod (left) following a speaker's self-initiated, same-turn, self-repair (Extract 1, lines 6–9; see also Supplementary Material 4 and 5 for video examples). Note that participants were facing each other in the laboratory and that these two images were taken from two frontal recordings (see Supplementary Material 1 for an illustration of participants' positioning in the laboratory).

In responding to this question, B produces two self-initiated, same-turn, self-repairs—each specifying some aspect of the respective prior unit. His first self-repair targets the type of agent (middle, line 3) he thinks is being administered by specifying it in line 4 (TMS of zo iets). The addressee registers this specification with a vocal continuer in line 5 (Ja). When B continues his response, resulting in a possible completion of a TCU in line 6, the addressee produces a vocal continuer (*mhm*; line 8) following a short gap (line 7). Note that the possible delay (line 7) may have influenced the speaker's production of the self-repair in line 9 and that the addressee's vocal continuer (line 8) occurs simultaneously with the onset of that self-repair. This second self-repair of the turn (line 9) again targets a referent mentioned in the prior unit as the repairable. In this case, the speaker provides a repair solution, specifying that it is not necessarily the brain as a whole that is being stimulated (as might be inferred from line 6) but only certain parts of the brain (*of bepaalde* *onderdelen van je hersenen*, line 9). The or-preface in this self-repair may already project that the trouble source formulation (line 6) is not about to be discarded as a whole but that merely a specification is coming instead (Lerner & Kitzinger, 2015, on or-prefaced self-repair in English). The addressee registers the repair solution with a long blink and a nod in line 9, before the speaker averts his gaze from the addressee and continues his telling in line 10.

The long addressee blink in response to disfluencies. In Extract 1, the long blink was combined with a nod. On the one hand, the co-occurrence of long blinks with nods supports the argument that long blinks serve a similar function. On the other hand, it raises the question whether long blinks per se can have a continuer function. Do long blinks alone also occur in positions where a continuer is relevant? They do, indeed. Long blinks alone were especially observed in response to intra-TCU troubles in speaking, that is, in response to disfluencies (e.g., sound stretches, cut-offs, *uh*, *uhm*; Clark & Fox Tree, 2002; Schegloff, 2010; Lickley, 2015) or to what Levelt (1989) called the "editing phase" in self-repairs—the interval between the end of a repairable and the beginning of a repair solution[10]. Previous research has already demonstrated that addressees responses (nods in this case) are especially relevant at points where the speaker has difficulties in producing a turn (Healey et al., 2013). In line with this research, Extract 2 demonstrates that continuers in general (in this case a vocal continuer and a long blink) are relevant in response to troubles in speaking, while Extract 3 exemplifies a case where a long blink alone is produced in this position. Prior to Extract 2, A and B talked about an upcoming pool party.

```
Extract 2 (ETC_19_582999)

01 A: Wat is nou-wat is nou het thema van-van-van de pool party
      what is now what is now the theme of of of the pool party
      What is the theme of the pool party now?

      ((Lines omitted))
```

[10] Four cases of long blinks in isolation also occurred at points of possible but not actual sequence completion—which would further support the argument that they can serve a continuer function (Schegloff, 1982, p. 84). However, since most of these were accompanied by laughter or larger gaze shifts—which could alternatively explain the eye closures (Evinger et al., 1994; see also Rossano, 2012; on sequence-final gaze withdrawal)—I did not include them in this analysis.

```
02 B: Wat ik had bedacht,
      what I had invented
      What I had invented

03   en dat vonden ze allemaal leuk,
     and that found they all lovely
     and they all liked the idea

04   is een Finding Nemo thema.
      is a   Finding Nemo theme
      is a   Finding Nemo theme

      ((Lines omitted))

05 B  en dan uh: (0.2)·[(0.2)+dumpen we]· ergen+s een visje,
      and then uh             dump  we somewhere a little fish
      and then uh              we dump a little fish somewhere
                            +averted gaze+
   a                   [mhm            ]
→                     ·●●●●●             .
```

In response to A’s question regarding the theme of the upcoming pool party (line 1), B answers by mentioning the “Finding Nemo” theme he had come up with (lines 2–4). When B has trouble producing an elaboration on the details of this theme as reflected by disfluencies (*en dan* followed by uh and a pause; line 5), the addressee produces a vocal continuer (*mhm*) and a long blink. This illustrates that continuers in general are not only relevant at TCU ends but that they can also be relevant within a TCU-in-progress, in response to disfluency-related disruptions of progressivity.

Similarly, Extract 3 shows an example of a case where a long blink alone is produced in response to a disfluency-related disruptions of progressivity within a TCU-in-progress. Prior to Extract 4, A and B have talked about possible romantic matches among friends.

```
Extract 3 (ETC_10_72939)

01 A: Maar wat zou je doen als je zelf in di-die positie zit=
      but what would you do if you self in thi this position sit
      but what would you do if you were in that position

02    =dan ga je dat ook wel doen
      then go you that  also  do
      then you would also do it

      ((Lines omitted))
```

```
03 B: En die zei ja ik v- ik denk da-z-prec-uh:-·geschikt ty·pe
      and he said yes I f I think tha sh prec uh suitable type
      And he said well I f I think tha sh prec uh is the right type
 → a                                                  ·●●●●

04    voor Helen is
      for  Helen is
      for Helen
```

In response to A's question (line 1), B reports that one friend had stated that another friend would be a suitable type for Helen (lines 3–4). After B produces several disfluencies in succession (multiple cutoffs plus uh: *ik v- ik denk da-z- prec-uh*, line 3), A produces a long blink alone. The succession of disfluencies in line 3 starts with a cut-off *ik v-* (recognizable as the beginning of the word *vind* 'find'), which is replaced with *ik denk*. Subsequently, the cutoff *da-z-* (potentially recognizable as the beginning of the words *dat* 'that' and *ze* 'she' respectively) is followed by another cut-off *prec-uh* (potentially recognizable as the beginning of the word *precies* 'exactly'), which is then replaced with *geschikt* 'suitable.' Note that the long addressee blink in this case is not produced in response to a repair solution (as in Extract 1), but it is produced in response to the disfluent editing phase (immediately after uh; line 3), and it only partially overlaps with the beginning of a repair solution (geschikt type, line 3) in this "replacing" type of self-initiated, self-repair (Schegloff, 2013).

Summary and discussion. First, we have seen that long addressee blinks were especially produced in response to repair solutions in speakers' self-initiated, same-turn, self-repairs. The long addressee blink in these positions (combined with a nod in 83% of cases) appears to pass up the opportunity to initiate repair or to take a full turn (Schegloff, 1982). By orienting to the continuation of a turn-in-progress, the addressee's long blink also seems to signal that by having received a repair solution the speaker has reached sufficient informativeness for current purposes. It may signal to the speaker that there is no need to clarify or specify further by producing more self-repairs, thereby helping the speaker in avoiding underinforming and overinforming (Grice, 1975; Mazeland, 2007; Sacks & Schegloff, 1979). If this is true, then speakers may strategically produce self-repairs in order to mobilize addressee responses (see Goodwin, 1980; Jefferson, 1974; Stivers & Rossano, 2010).

Furthermore, we have also seen that long blinks often co-occurred with other continuers, which raises the question whether long blinks contribute anything at all in these cases and whether, for example, blinks are not a side effect of nodding. Since head movements often co-occur with gaze shifts and gaze shifts often co-occur with blinks (Evinger et al., 1994), blinks may simply be a way to optimize vision while nodding. Evinger et al. (1994) stated that, “since saccadic suppression impairs vision during a saccadic gaze shift, a gaze-evoked blink can lubricate the cornea without interfering with vision” (p. 342). However, there are at least two reasons to doubt such an account. First, in the data most blinks did not co-occur with nods (see Figure 5), and 36% (n = 30) of all nods (n = 83) were not combined with blinks at all. Secondly, addressees typically do not shift gaze while nodding; they tend to keep looking at the speaker’s face. Yet, while blinks may not be caused by nods, there may be motoric synergies between producing a nod and producing a blink (e.g., the downward movement of the head facilitating the downward movement of the upper eyelid).

Secondly, we have seen that long blinks alone (i.e., combined with neither a nod nor a vocal continuer in 75% of cases) were also used in positions where a continuer is relevant, namely, in response to intra-TCU disfluency-related disruptions of progressivity. What might be the function of a long blink produced in this position? First, producing a long blink in this position may display continued recipiency. Like TCU ends, intra-TCU disfluencies may be vulnerable places at which addressees might steal the turn (Jefferson, 1983; Maclay & Osgood, 1959). By producing a long blink, the addressee may display attentiveness through responsiveness, passing up the opportunity to take a turn and thereby aligning with the telling in progress. Secondly, it has been argued that speakers prosodically distract their addressee’s attention from the repairable in their self-repairs by increasing the pitch and loudness of the following repair proper (Noteboom & Quené, 2014). Similarly, long blinks (as the ones in Extracts 2–3) may contribute to camouflaging or rendering invisible the presumably unintended reduction in progressivity, thereby minimizing the potentially face-threatening act (Brown & Levinson, 1987) of “watching” the speaker having difficulties speaking. On a related note, research has shown that perceived direct gaze is cognitively demanding (Conty, Gimmig, Belletier, George, & Huguet, 2010). If this is true, then addressees may produce long blinks during disfluencies to momentarily reduce the speaker’s cognitive load. By blinking, the addressee disrupts eye contact, which might help the speaker make her speech fluent again—orienting to progressivity while minimizing "joint effort" (Clark, 1996). In short, in the face of disfluencies, then, long addressee blinks are presumably used to display continued recipiency, to camouflage the disfluencies, to reduce the speaker's cognitive load, or are due to a combination of these potential functions.

Finally, while the long addressee blink cases presented here already provide suggestive evidence, future research will be required to provide conclusive evidence. More conversation analytic research would be desirable to further demonstrate that participants orient to long blinks as consequential social actions, while experimental research will be required to confirm the hypothesis that speakers are causally influenced by addressee blink behavior. Based on the present study, however, we would like to suggest that since long addressee blinks appeared in the same sequential positions as continuers, they are likely to serve similar functions: orienting to progressivity by passing up the opportunity to take a turn or to initiate repair (Robinson, 2014; Schegloff, 1982) and aligning with the telling-in-progress (Stivers, 2008) while signaling successful grounding (Clark, 1996).

General discussion

Does addressee blinking serve a feedback function in conversation? First, the present results are incompatible with a purely physiological interpretation of blinking in conversation. While all blinks also lubricate the cornea, addressees blinked too frequently (approximately 30 times per minute) to serve solely this physiological function (Doane, 1980). More strikingly, these results are consistent with a social-communicative interpretation of blinking because addressee blinks were timed around TCU ends, the location where addressee responses are typically produced and where speakers tend to visually monitor addressees for feedback (Bavelas et al., 2002).

Second, short and long addressee blinks appear to fulfill partially different functions. While there was no clear difference in timing between short and long blinks relative to TCU ends, long blinks—in contrast to short blinks—were (a) less frequent; (b) more likely to co-occur with nods or vocal continuers; (c) more likely to occur during mutual gaze; and (d) their production was restricted to sequential contexts in which it was structurally relevant to display recipiency and understanding, together suggesting that long blinks are particularly likely to serve a social-communicative function— especially when combined with other feedback responses.

Can a cognitive interpretation of the timing of addressee blinks be ruled out? The fact that the majority of addressee blinks were timed to TCU ends is also consistent with a cognitive interpretation of blinking (Nakano et al., 2009, 2013; Siegle et al., 2008). Nonblinking while listening to the core of a TCU may be a symptom of addressees' high cognitive load, while blinking at the end of a TCU may reflect addressees' relative decrease in cognitive load. While this cognitive interpretation may hold for short blinks, it seems less plausible for long blinks. Short blinks occurred throughout whole conversations. Long blinks, however, were placed in specific sequential positions and, although this is an open issue requiring further investigation, it seems unlikely that in these positions addressees' cognitive load was particularly low relative to other sequential positions.

Cognitive, perceptual, and social functions of addressee blinks are not mutually exclusive. It is possible that the cognitive and perceptual functions underlie and evolutionarily preceded any potential social-communicative function. Perhaps blinking as a symptom of low cognitive load or attentional disengagement and the need to control blinking to minimize audiovisual information loss during speech comprehension (McGurk & MacDonald, 1976; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007) have been co-opted for communicative purposes, such that they are now used as a semiotic signal. As squinting the eyes—as if to see more clearly—seems to signal lack of understanding, closing the eyes by blinking seems to convey "no need to see anymore" because understanding has been established (Lakoff & Johnson, 1999, on the Understanding-Is-Seeing metaphor).

Interestingly, the results suggesting a continuer function for addressee blinking in Dutch are in line with studies on blinking in Yélî Dnye (Levinson, 2015; Levinson & Brown, 2004) and American Sign Language (Sultan, 2004). At least based on this limited number of studies, addressee blinks as signals of successful grounding seem to be independent from language modality—since it is used in spoken as well as signed language—as well as from language history—since it has also been described in Yélî Dnye, a Papuan language. If addressee blinking as a signal of recipiency and successful grounding is shared across a wider range of unrelated languages, it may have evolved due to common pressures of a shared conversational infrastructure (Dingemanse, Torreira, & Enfield, 2013; Stivers et al., 2009).

Acknowledgements

We thank Binyam Gebrekidan Gebre, Mart Lubbers, Emma Valtersson, and Sean Roberts for programming support, and Herb Clark, Lorenza Mondada, the Language & Cognition Department, and members of the Dialogue Project at the MPI Nijmegen for valuable discussions.

Appendix

Transcription conventions (based on Mondada, 2014)

Gestures and descriptions of embodied actions are delimited between two identical symbols and are synchronized with correspondent stretches of talk (one symbol per embodied action: × for addressee blinks [each ● reflects approximately 100 ms], * for nods, + for speaker gaze aversion during the telling, ∞ for addressee gaze aversion during the telling,♦for other bodily conduct). The action described continues until after the excerpt's end (*--->>) or across subsequent lines (*--->) until the same symbol is reached (---->*). Participant are identified in the margins. Capital letters (e.g., A) indicate speakers, small letters (e.g., a) indicate addressees. The moment at which a still image has been taken is indicated with a # showing its position within the turn. The corresponding figure number is shown in a separate line (e.g., #fig. 6).

Chapter 3. Eye blinks are perceived as communicative signals in face-to-face interaction[11]

Abstract

In face-to-face communication, recurring intervals of mutual gaze allow listeners to provide speakers with visual feedback (e.g. nodding). Here, we investigate the potential feedback function of one of the subtlest of human movements—eye blinking. While blinking tends to be subliminal, the significance of mutual gaze in human interaction raises the question whether the interruption of mutual gaze through blinking may also be communicative. To answer this question, we developed a novel, virtual reality-based experimental paradigm, which enabled us to selectively manipulate blinking in a virtual listener, creating small differences in blink duration resulting in 'short' (208 ms) and 'long' (607 ms) blinks. We found that speakers unconsciously took into account the subtle differences in listeners' blink duration, producing substantially shorter answers in response to long listener blinks. These findings suggest that, in addition to physiological, perceptual and cognitive functions, listener blinks also serve communicative functions, directly influencing speakers' communicative behavior in face-to-face communication. More generally, these findings may shed new light on the evolutionary origins of mental-state signaling, which is a crucial ingredient for achieving mutual understanding in everyday social interaction.

[11] Hömke, P., Holler, J., & Levinson, S. C. (in press). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One*. doi: 10.1371/journal.pone.0208030.

Introduction

Human communication is a joint activity (Clark, 1996). Rather than just one party being active at a time by producing speech, both speaker *and listener* contribute signals critical to progressing the exchange of information. Listener feedback (or 'back-channel' responses; Yngve, 1970), such as *mhm* or *uhu,* are crucial for successful communication since they facilitate the process of grounding (Clark & Brennan, 1991) and thus the achievement of mutual understanding. Eliminating or reducing listener feedback is detrimental to speakers' behavior (Bavelas, Coates, & Johnson, 2000).

Unlike other animals, humans tend to engage in mutual gaze when communicating without necessarily signaling aggressive intent or affiliative interest (Rossano, 2013; Levinson & Holler, 2014). For successful face-to-face communication, these recurring intervals of mutual gaze are important, as they allow listeners to also provide speakers with *visual* feedback, such as nodding. Here, we investigate a behavior that—unlike nodding—is not commonly known for its communicative function, or for its role in the process of grounding: eye blinking.

Infants hardly blink (Ponder & Kennedy, 1927), but blink rate increases over time until adulthood (Zametkin, Stevens, & Pittman, 1979). Adults blink more frequently than physiologically required for ocular lubrication (Doane, 1980). We blink approximately 13,500 times every day—thus making it the most frequent facial action—with blinks being among the fastest movements the human body can generate (Peshori, Schicatano, Gopalaswamy, Sahay, & Evinger, 2010). In addition to reflex protective and physiological eye-wetting functions (Doane, 1980), blinking has been shown to index cognitive load. Under low cognitive load, people blink more than under high cognitive load (Ponder & Kennedy, 1927; Siegle, Ichikawa, & Steinhauer, 2008). A neuroimaging study has corroborated these behavioral findings showing that blinking deactivates the dorsal attention network while activating the default-mode network, suggesting an active involvement of blinking in attentional disengagement (Nakano, Kato, Morito, Itoi, & Kitazawa, 2013). Blinking has also been linked to social functions. Comparing different species of non-human primates has revealed that blink rate correlates positively with social group size (Tada, Omori, Hirokawa, Ohira, & Tomonaga, 2013), a measure of social complexity associated with neocortex size (evidence used to support the "social brain hypothesis"; Dunbar, 1992). Further, comparing different activity types in humans—that is, looking at a still target, reading, and conversation—the highest blink rate was found in conversation (Doughty, 2001). This suggests that in conversation blinking may fulfill functions beyond physiological and cognitive functions.

A recently published study corroborates this suggestion by taking a close look at the statistical distribution of eye blinking in face-to-face conversation (Hömke, Holler, & Levinson, 2017). While our intuitive assumption may be that eye blinking is something we do at entirely random points while we listen to someone else speak, the study revealed the opposite to be the case: the majority of listener eye blinks occurred in typical 'feedback slots' during conversation, namely around the end of speaking units (or turn-constructional units, i.e., junctures at which a speaking turn may be perceived as possibly complete; Sacks, Schegloff, & Jefferson, 1974). Voluntary blinks have longer durations (Kaneko & Sakamoto, 1999) and longer blinks have been suggested to have a special communicative salience (Levinson & Brown, 2004; Mandel, Helokunnas, Pihko, & Hari, 2014). Hömke et al. (2017) therefore categorized blinks into short and long listener blinks, based on the distribution of blink durations observed in their corpus of face-to-face conversations, using a threshold of 410 ms, separating the longest 25% from the rest. In contrast with short blinks, long blinks (see S1 Video for an example) were more likely to be produced during the mutual gaze window (Bavelas et al., 2002), in co-occurrence with other listener feedback responses like nods, and they were produced in specific communicative contexts in which signaling understanding was especially relevant. Together, these findings suggest that even subtle movements such as eye blinks may be perceived as meaningful signals by others. If this is indeed so, then we should be able to pinpoint some direct communicative consequences that result from the perception of listeners' eye blinks. However, the corpus data observed by Hömke et al. (2017) is limited in the extent to which it can provide such evidence due to the correlational nature of corpus studies. Here, we build on this earlier research by experimentally testing whether listener blink behavior has any measurable effect on speakers' speech production.

Questions regarding the causal role of subtle facial cues in interactive face-to-face communication have previously been impossible to address with a high degree of experimental control. Hence, we developed a novel experimental paradigm using Virtual Reality technology enabling us to selectively manipulate blink duration in a virtual listener. Participants were asked to have a conversation with three different avatars and to respond to open questions (e.g., *How was your weekend, what did you do?*). While participants were answering, the avatar produced different types of visual feedback, which was triggered secretly by a confederate, who could see and hear the participant (via a video-camera link) but who was blind to the conditions and hypotheses and instructed to press a button whenever it felt appropriate to signal understanding. The confederate's button presses triggered either (1) no listener feedback in the avatar (control condition), (2) nods with short blinks or (3) the same nods but with slightly longer blinks. The nods accompanied the avatar's blinking behavior to mimic the typical natural environment of blinks that occur in feedback slots in conversation (Hömke et al., 2017). Crucially, the nods in the two conditions were identical in form and duration such that the only difference between the two conditions was the duration of the co-occurring blinks.

The temporal length of participants' answers was compared across conditions from the first to the last vocalization produced by the speaker in response to each question. The rationale was that if listener blink duration is irrelevant for the speaker's speaking behavior, one should not observe any differences in answer length between the nod with short blink and the nod with long blink condition. Alternatively, if listener blinking can indeed function communicatively, then differences in speaker behavior would be expected. Specifically, if nods with long blinks indeed function as a "move on" signal of understanding, signaling "I've received enough information for current purposes", as has been suggested by Hömke et al. (2017), answers should be shorter in the nod with long blink than in the nod with short blink condition.

Speaking behavior, like any other social behavior, varies from individual to individual (Heerey, 2015). One particular individual difference measure of dispositional social sensitivity—the Empathy Quotient (Baron-Cohen & Wheelwright, 2004)—may modulate the perception of eye blinks. That is, sensitivity to blink feedback may depend on the speaker's degree of empathy, which is the "drive or ability to attribute mental states to another person/animal, and entails an appropriate affective response in the observer to the other person's mental state" (Baron-Cohen & Wheelwright, 2004, p.168). It has been observed that "to drive your point home in a discussion for far longer than is sensitive to your listener" constitutes low-empathy behavior (Baron-Cohen & Wheelwright, 2004, p.170), suggesting that high-empathy speakers may be more sensitive to listener feedback in general. More specifically, it has been shown that, passively watching another person telling a story on video, brains of high-empathizers were more responsive (i.e., showed a higher amplitude) to the narrator's spontaneous eye blinks than brains of low-empathizers (the measured brain responses to observed blinks were magnetoencephalographic and they peaked at about 250 ms in the parieto-occipital cortex; Mandel, Helokunnas, Pihko, & Hari, 2015). Interestingly, the brain response to observed blinks was only modulated by empathy when the video of the person telling the story was presented with sound, "most likely because the human voice created a social context that enforced the empathy-related modulation on brain activity" (Mandel et al., 2015, p.15). These findings suggest that high-empathizers may be especially responsive to (variations of) blink behavior in face-to-face communication. To address this issue, the random sample of healthy participants in this study was asked to complete the Empathy Quotient questionnaire (Baron-Cohen & Wheelwright, 2004) after the experiment.

The overall aim of the current study was to experimentally test earlier claims based on correlational evidence suggesting that listener blinking may serve a communicative function in conversation (Hömke et al., 2017). The hypothesis was that speakers would produce shorter answers when talking to a listener providing feedback in the form of nods with long instead of short blinks (and that this effect may be modulated by speakers' empathy). The present study demonstrates, for the first time, a sensitivity of speakers to listener blink behavior as a communicative signal in interactive face-to-face communication.

Materials and methods

Participants

Forty native Dutch speakers were recruited through the MPI for Psycholinguistics - subject database (www.mpi.nl/ppreg) to participate in the experiment. Data of four participants could not be used due to technical error. Data of one additional participant were excluded from all analyses because he excessively looked away from the screen (more often than 2.5 SD above the mean) during avatar listener responses (and therefore could not have been influenced by them), resulting in a final sample of 35 participants (18-38 years; mean age = 21.88; 19 females, 16 males). The sample size was planned in accordance with prior studies on listener feedback (Bavelas, Coates, & Johnson, 2002). The whole session lasted about sixty minutes and each participant was paid €10.

Apparatus and materials

Laboratory set-up and equipment. The experiment took place in the Virtual Reality laboratory at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. Participants were seated in front of a computer screen (HP Compaq LA2405WG) with speakers (Hercules XPS 2.010) wearing a lightweight, head-mounted microphone (DPA-d:fine-88). Three synchronized video cameras (Sony 3CCD Megapixel) were used to record the participants (1) frontally, and (2) laterally, as well as to record a separate computer screen showing precisely what the participant was seeing on their screen (i.e., the avatar), thus allowing us to temporally link participant and avatar behavior (see Fig.1). Audio was recorded using Adobe Audition CS6. Thus, each recording session resulted in three videos and one audio file, which were synchronized based on audible and visible markers (produced at the beginning of each block) and exported in Adobe Premier Pro CS6 (MP4, 25 fps).

Figure 1. Virtual listener (left) interacting with human speaker (right) in the present experimental set-up (see also S2 Video).

In the control room next to the experiment room, a confederate (see Procedure) was seated in front of a keyboard (Apple MB110LL/B) and a computer screen (17"Acer AL732) showing the participant from a frontal view in real time. Audio from the participant's head-mounted microphone was also directly transmitted to the control room and played via speakers (Alesis M1Active 520) in real time.

Avatar characteristics and behaviors. The experiment was programmed using

WorldViz's Vizard 5.5. Three different female avatars were created based on a stock avatar produced by WorldViz. The avatars' speech was pre-recorded by three different female native speakers of Dutch and played at relevant points during the experiment. The avatars' lip movements were programmed to match the amplitude of the pre-recorded speech files, creating the illusion of synchronization. The speech materials consisted of a general introduction (e.g., *Hoi, Ik ben Julia, leuk je te ontmoeten!*; 'Hi, I'm Julia, nice to meet you!' and *Ik heb een aantal vraagen aan jou*; 'I have a couple of questions for you') and a set of 18 open-ended questions (e.g., *Hoe was je weekend, wat heb je allemaal gedaan?*; 'How was your weekend, what did you do?'). The avatar also gave a response to the participant's answer (e.g., *Oh ja, wat interessant!*; 'Oh, how interesting!') before proceeding to the next open question (e.g., *Ik heb nog een vraag aan jou;* 'I have another question for you'), or to close the interaction (*Hartelijk bedankt voor dit gesprek, ik vond het gezellig!*; 'Thank you very much for this conversation, I enjoyed it!') (see S2 Video for an example of a trial). Whenever an avatar was in the listener role, her feedback responses—the crucial experimental manipulation in the present study—were modeled on typical feedback behavior that occurs in natural conversation. These behaviors consisted of (1) a head nod accompanied by a short blink (208 ms from blink onset to blink offset) and (2) the same head nod but accompanied by a longer blink (607 ms from blink onset to blink offset). The durations were based on the average durations of short and long blinks in a corpus of Dutch face-to-face conversations (Hömke et al., 2017) and the nods (duration: 499 ms from nod onset to nod offset) accompanied the avatar's blinking behavior to mimic the typical natural environment of blinks occurring in feedback slots in conversation (ibid.). Since in the corpus data, the onsets of nods and blinks typically coincided (perhaps due to motoric synergies, i.e., the downward movement of the head potentially facilitating the downward movement of the upper eyelid; Hömke et al., 2017, p.12), the onsets of nods and blinks in the current study were programmed to coincide as well. Since the offsets of nods and blinks were relatively varied or uncoordinated in the corpus data, the offsets of nods and blinks in the current study were programmed in a way that impressionistically achieved the most natural look, resulting in long blinks lasting about 100 ms longer than the nods (see exact durations above). Crucially, the nods in the two conditions were identical in form and timing such that the only difference between the two conditions was the duration of the co-occurring blinks.

Design

A within-subject design was used with listener feedback (none, nod with short blink, nod with long blink) as independent variable, mean answer length as the dependent variable, and Empathy Quotient as an individual difference measure. The experiment consisted of three blocks, one block per feedback condition and avatar. The set of 18 question stimuli were split up into three sets of 6 questions and each set was assigned to one of the three avatars, meaning each participant heard each question only once. The order of feedback conditions as well as the assignment of avatars (and thus the 6 questions that were paired with the respective avatars) to the listener feedback conditions was counterbalanced across participants. The order of items (question stimuli) within each block was randomized.

Procedure

On arrival in the laboratory, participants were seated in front of the computer screen (see Fig. 1) and were asked to meet and have a conversation with three different avatars and to respond to their questions. After a short personal introduction, the avatar asked questions and produced different types of visual feedback responses while participants answered (see Avatar characteristics and behavior section). The visual feedback responses were triggered secretly by a confederate, a Dutch native speaker who could see and hear the participant (via a video-camera link) but who could not see the avatar and the feedback behaviors being triggered. The confederate was blind to the experimental hypotheses and not informed about the manipulations.

The confederate was simply instructed to press a button whenever it felt appropriate to signal understanding while listening to the participants' answers. Upon each answer completion by the participant, the avatar produced a response to the participant's answer (e.g. 'Oh, how interesting!'), which was also triggered secretly by the confederate. After having finished the conversation with the third avatar, the experiment was over and participants were asked to complete questionnaires (see above) before they were debriefed on the purpose of the experiment. Written informed consent was obtained before and after the experiment. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen. Written informed consent was obtained from all participants visible in still images or video footage included as part of this manuscript to publish the respective case details.

Measures

Behavior Coding. Answer length, the dependent variable in this study, was measured from the first vocalization (excluding in-breaths) to the last vocalization produced by each speaker in response to each question. Answers were always embedded between the offset of the avatar's question and the onset of the avatar's uptake, determined by when the confederate pressed the "Oh, how interesting!" button. Speech disfluencies such as "um" and "uh" were treated as being part of the answer. Answer lengths were annotated in ELAN 4.9.3 (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006).

Questionnaires. Two questionnaires were used. Firstly, we used the Dutch version of the Empathy Quotient questionnaire first developed by Baron Cohen and colleagues, a widely used questionnaire measuring both cognitive aspects of empathy (i.e., understanding and/or predicting what someone else might think, feel, or do) and affective aspects of empathy (i.e., feeling an appropriate emotion triggered by seeing of another's emotion; test-retest reliability: $r = 0.97$, $p < .001$, as reported by Baron-Cohen & Wheelwright, 2004). It consists of 60 statements (e.g., *I find it easy to put myself in somebody else's shoes* or *I can pick up quickly if someone says one thing but means another*) and participants indicate on a four-point scale to what extent they agree with each statement (strongly agree, slightly agree, slightly disagree, strongly disagree). Secondly, we used a questionnaire assessing any explicit awareness of the different feedback types, that is, whether participants had noticed nodding and/or blinking in the virtual listeners at all, and if so, if they had noticed any variation in these behaviors across conditions.

Statistical analysis

The answer length data set was trimmed for outliers deviating more than 2.5 standard deviations from the mean answer length (excluding 3% of the entire dataset; these were often cases where speakers started vocalizing, e.g. '*uuhhm*', but then hesitated for a very long time thinking about what to answer). We used R (Team RC, 2014) and *lme4* (Bates, Maechler, Bolker, & Walker, 2014) to test in a linear mixed-effects model whether answer length differed depending on listener feedback. The initial model was an intercept-only model estimating the mean answer length including intercepts for items (question stimuli) and participants as random effects. Using a likelihood ratio test (using the ‘anova’ function), this intercept model was compared to a model, which differed only in that listener feedback (no feedback, nods with short blink, nods with long blink) was included as a fixed effect. To test whether any effect of listener feedback on answer length was modulated by the speakers’ empathy, we first entered listener feedback (no feedback, nod with short blink, nod with long blink) and speaker empathy (EQ score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items (question stimuli) and participants as random effects into the model. This model was then compared to a model that only differed in that listener feedback and speaker empathy were entered as fixed effects *with* interaction term, again using a likelihood ratio test (with the ‘anova’ function).

Results

Were speakers sensitive to listener blink duration in face-to-face communication? The initial model was an intercept-only model estimating the mean answer length in seconds, including intercepts for items (question stimuli) and participants as random effects (β = 40.16, SE = 3.1, t = 12.95). Adding listener feedback as a fixed effect improved the model fit significantly ($\chi^2(2) = 7.56$, p = .022), revealing that, relative to talking to a listener providing feedback in the form of nods with short blinks (β = 40.82, SE = 1.68), speakers indeed produced shorter answers when talking to a listener providing feedback in the form of nods with long blinks (β = -2.86, SE = 1.42, t = -2.016, p = .044). Relative to talking to a listener providing no feedback (β = 41.7, SE = 3.2), speakers also produced shorter answers when talking to a listener providing feedback in the form of nods with long blinks (β = -3.74, SE = 1.41, t = -2.641 p = .008). Speakers’ answer length in the ‘nod with short blink’ condition and the ‘no feedback’ control condition were statistically indistinguishable (β = 0.87, SE = 1.41, t = 0.618, p = .537). These results show that speakers were not only sensitive to the absence of visual feedback, but also to subtle differences in blink duration (see Fig.2)[12].

[12] Predicting confederate’s feedback button press frequency (number of button presses per answer divided by the length of the same answer in minutes; M = 7.4; SD = 2.6) by feedback

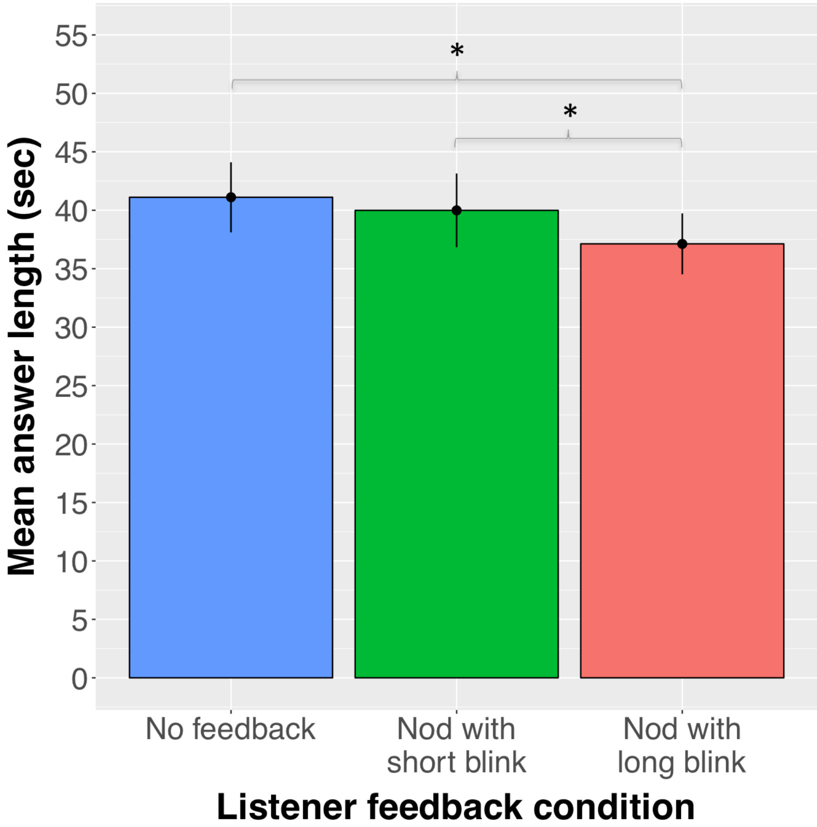

Figure 2. Mean answer length (sec) by listener feedback. Standard errors are represented in the figure by the error bars attached to each column.

Was speakers' sensitivity to listener blink duration affected by their degree of empathy? We predicted answer length by entering listener feedback (no feedback, nod with short blink, nod with long blink) and speaker empathy (M = 39.56; SD = 10.66) as fixed effects (without interaction term), and intercepts for items (question stimuli) and participants as random effects into the model. This model was compared to a model that only differed in that listener feedback and speaker empathy was entered as fixed effects *with* interaction term. Including listener feedback and speaker empathy *with* interaction term did not improve the model fit significantly ($\chi^2(2)$ = 3.808, p = .148). These results show that the effect of blink duration on answer length was unaffected by the speakers degree of empathy. The fact that not only high-empathizers but speakers in general were sensitive to differences in blink duration is

---

condition (nod with short vs. long blink), including random intercepts for participants and items, confirmed that button press frequency was consistent across conditions (β = 2.83, SE = 3.54, t = 0.801, p = .424).

in line with the hypothesis that listener blinking can indeed be perceived as a communicative signal.

Were speakers aware of the difference between short and long listener blinks? To address this question, participants were asked to complete a post-experiment questionnaire assessing whether they had noticed listener nodding or blinking. While all participants noticed the listeners' nodding, only about half (46%) of the participants noticed the listeners' blinking. Most importantly, none of the participants reported having noticed any variation in blink behavior across conditions, suggesting a covert effect of listener blink duration on speakers' utterances.

Discussion

Are speakers sensitive to listener blink duration in face-to-face communication? The results reveal that subtle, millisecond differences in blink duration caused speakers to design answers to questions that differed considerably in length, namely by several seconds. Speakers produced shorter answers when talking to a listener providing feedback in the form of nods with long blinks instead of short blinks. Despite the striking effect on the speakers' linguistic behavior, listeners' blink behavior appeared to escape speakers' explicit awareness.

These findings have theoretical implications. Although natural human language is multimodal and social-interactive in nature, traditional models of language processing have primarily focused on verbal language and on utterances produced outside of a social-interactive context. This study embraces the multimodal as well as the social-interactive nature of language and it may provide further motivation for a paradigm shift, an 'interactive turn' (Kendrick, 2017, p.7) that is already taking place in psycholinguistics (Pickering & Garrod, 2004; Levinson, 2016; Holler, Kendrick, & Levinson, 2017), but also in the cognitive sciences more generally (Sebanz, Bekkering, & Knoblich, 2006; Jaegher & Paolo, 2010; Schilbach et al., 2013; Fröhlich et al., 2016). More specifically, the finding suggesting that long listener blinks may be functionally involved in managing mutual understanding highlights that speaking in interaction is not a unilateral process but a joint activity involving active contributions from both speaker *and* listener (Clark, 1996; Yngve, 1970; Bavelas et al., 2000). It further emphasizes that speaking not only involves self-monitoring (Levelt, 1983) but also other-monitoring (Clark & Krych, 2004). As Levelt (1989, p.8) put it, "A speaker, while delivering his utterance, is continuously monitoring himself and his interlocutors, and this feeds back to what he is doing". He further notes that "interlocutors send various signals to the speaker which tell him that (...) he should go on" and "much of this can be done by gaze or gesture". The long listener blink as described in the present study appears to constitute such a gesture, a facial gesture that may be perceived as signaling "you've been sufficiently informative". As such, the long listener blink is a type of backchannel (Yngve, 1970). More specifically, it is what Bavelas and colleagues (2014, p.16) called a "facial backchannel", a facial listener response providing "rapid feedback to the speaker without interrupting or taking up a turn".

In the following, we discuss a number of open questions, limitations, and potential avenues for future research. So far, we have suggested that long listener blinks are perceived as a communicative signal of understanding. However, one may wonder whether the long listener blink is not merely a symptom (e.g., of low cognitive load) that is informative for the speaker but that is not communicatively intended by the listener (Grice, 1957; Brennan, Galati, & Kuhlen, p.314). While this study does not address this issue conclusively, there are a number of reasons to believe that the long listener blink is indeed a communicatively intended signal of understanding. First, the longer duration of a long listener blink is an indicator of a voluntarily produced blink as opposed to a spontaneously produced blink, which is characterized by shorter durations (Kaneko & Sakamoto, 1999). Secondly, analyses of listener blinks in natural conversation (Hömke et al., 2017) revealed that in contrast with short blinks, long blinks were more likely to be produced during the mutual gaze window (Bavelas et al., 2002) and they were produced in specific communicative contexts in which signaling understanding was especially relevant. Together, this suggests that long blinks may indeed be communicatively intended and specifically designed to signal 'I understand, I've received sufficient information' (Clark & Schaefer, 1989). However, because this experiment was designed to test the consequences of perceiving listener signals on speaker behavior rather than the production of listener signals, we cannot be sure that long blinks are indeed used by listeners in this way. For example, one alternative possibility is that a nod with a long blink was simply interpreted as a signal to stop speaking, whether or not the answer had been understood. Thus, future experimental work is required to provide conclusive insights into the extent to which long listener blinks are communicatively intended, and what specifically they are meant to convey.

Another question that emerges from these findings is why speakers' answer length did not differ when talking to a listener not providing any feedback (no feedback control condition) as compared to when talking to a listener providing feedback in the form of nods with short blinks. The nods with short blinks condition in this study was designed to be the 'unmarked' baseline condition, intended to signal that the listener currently has no special informational needs, i.e., that the speaker is neither under-telling nor over-telling (Grice, 1975; Levinson, 2000; Schegloff, 2007, p.133). The observation that absence of this sort of feedback does not affect speaker behavior seems at odds with previous research demonstrating the importance of listener feedback for speaking in face-to-face communication (Bavelas et al., 2000). Note, however, that in that prior study, speakers interacted with a human listener who was distracted by a secondary cognitive task, resulting in deviant or absent listener responses at points where they were expected. In the current experiment, by contrast, speakers interacted with a virtual listener who did not provide any listener responses at all in the 'no feedback' condition. Presumably, if speakers did not get any feedback while answering the first question in the 'no feedback' condition, they may have assumed that this avatar is not "capable" of giving feedback at all, thus the speaker may not have expected any listener responses and produced his or her answers accordingly (as if talking to an answering machine). In fact, previous research has shown that expectations about the listener's feedback can change the effect of listener feedback (or the omission thereof) (Kuhlen & Brennan, 2010). It thus seems plausible that the participants' expectations about the avatar's behavioral repertoire may explain the lack of a difference in answer length between the short-blink and the no-feedback control condition. However, we should also bear in mind that, despite showing no difference in answer length, the quality of the answers may well have differed (e.g., answers in the no feedback condition may have contained more hesitations instead of semantic content). To explore this possibility, a subset of the speaker response data (n = 108) was coded for filled and unfilled pauses. These initial analyses did not suggest any differences in response quality, at least not with respect to filled and unfilled pauses. Since no substantial numerical differences were observed across conditions, the coding and analysis was not followed up for the remainder of the data due to the time-consuming nature of the analyses.

Since the current study has focused on the relatively global measure of answer length, future studies may zoom in on more immediate, local effects listener feedback may have on the content (e.g., level of detail, information density) as well as on speech production (e.g., hesitations, speech rate). Interestingly, at least impressionistically, there were no such effects on answer production. Also, the systematic analysis of hesitations for a subset of the data supported this impression, since it did not suggest any differences in the global amount of hesitations produced across conditions (i.e., filled and unfilled pauses), which may at least suggest that the shorter answers in the nod with long blink condition were not the result of speakers perceiving the avatar's behavior as disruptive or interruptive.

On a related note, the main reason why speakers may have had varying expectations about the avatar's feedback behavioral repertoire is that, in the current study, listener feedback was manipulated between avatars. One important reason to manipulate listener feedback between avatars was that we wanted to include a no feedback control condition. Since this is a novel paradigm, we needed a way to know whether speakers take into account listener feedback produced by an avatar at all (in case we had not found any differences in answer length between the nod with short versus nod with long blink condition). Thus, we manipulated listener feedback between avatars because we were concerned that interacting with an avatar providing human-like feedback in one trial but then suddenly no feedback in the next trial (control condition), may be too confusing for participants, disrupting the natural flow of the whole interaction with that avatar. Now that the current study has established that speakers are indeed sensitive to avatar listener feedback in general, and more specifically, to listener blink behavior, future studies may zoom in further on the moment-by-moment dynamics of the possibility of managing mutual understanding through blink feedback. Specifically, they may consider manipulating listener blink duration within individual virtual listeners, since in natural conversation, the same listener may sometimes display short blinks and sometimes long blinks, depending on situationally shifting informational needs. At this point, we can only speculate on the outcome of such a change in paradigm, but a likely possibility is that the contrast effect of perceiving both short and long blinks within one avatar may enhance the pattern of the current findings.

Taken together, the findings indicate that even visually subtle behavior such as listener blinking is anything but irrelevant in face-to-face communication. The different functions of listener blinking are, of course, not mutually exclusive. Like eye gaze in social contexts (Rossano, 2012, p.312; Gobel, Kim, & Richardson, 2015) it appears that eye blinking may serve self-oriented and other-oriented functions at the same time. In addition to physiological, perceptual and cognitive functions, listener blinks may serve a crucial communicative function in face-to-face communication. If this is true, cognitive and perceptual functions very likely preceded the communicative function, phylogenetically as well as ontogenetically. Blinking as a consequence of low cognitive load or attentional disengagement, and the need to control blinking to minimize audio-visual information loss during speech comprehension (McGurk, 1976, Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2006) may have been co-opted for communicative purposes through processes of ritualization (Darwin, 1872; Tomasello & Call, 2007), which would suggest a non-arbitrary, iconic relationship between form and meaning in communicative blinking (Grice, 1957; Perniss & Vigliocco, 2014). Thus, the present findings may shed new light on the postulated "embodied" origin of the Understanding-Is-Seeing metaphor (Lakoff & Johnson, 1999), and more generally, on the visual origins of mental-state signaling (Lee, Mirza, Flanagan, & Anderson, 2014), a crucial ingredient for achieving intersubjectivity in everyday face-to-face communication. Moreover, they corroborate the notion that speaking in interaction is not a unilateral process but a joint activity involving active contributions from both speaker *and* listener (Yngve, 1970; Clark, 1996; Bavelas et al., 2000).

Acknowledgments

We thank the participants in this study, Jeroen Derks and Han Sloetjes for programming support, Renske Schilte who served as the confederate, and Herb Clark, David Peeters, Florian Hintz, Markus Ostarek, Wendy Sandler, the Language & Cognition Department and members of the Virtual Reality Focus Group at the MPI Nijmegen for valuable discussions. We also thank the Max Planck Gesellschaft and the European Research Council (Advanced Grant INTERACT #269484 awarded to S. C. Levinson) for financially supporting this research.

Supporting information (see https://pure.mpg.de/pubman/item/item_3008949)

S1 Video. Long listener blink. Example of a long listener blink as used in face-to-face conversation (Hömke, Holler, & Levinson, 2017).

S2 Video. Example of a trial (short blink). Example of a trial in the nod with short blink condition, including the avatar's question, the avatar's nods with short blinks during the participant's answer, and the avatar's response following answer completion.

S3 Video. Example of a trial (long blink). Example of a trial in the nod with long blink condition, including the avatar's question, the avatar's nods with long blinks during the participant's answer, and the avatar's response following answer completion.

S1 Data. Dataset underlying the findings.

S1 Text. Statistical model.

Signaling non-understanding facially

Chapter 4. Eyebrow movements as signals of communicative problems in face-to-face conversation

What is the role of the face in face-to-face social interaction? It is well known that the face plays an important role in expressing emotions (Darwin, 1872; Ekman, 1993). *Facial expressions* are often considered to be rather involuntary manifestations of an individual's emotion (e.g., fear upon seeing a spider) and they have been distinguished from more voluntary *facial gestures* (Bolinger, 1946, Kendon, 2004; see also "conversational facial signals", Ekman, 1979; "facial displays", Kraut & Johnston, 1979). Rather than being part of an individual-emotional process, facial gestures are considered to be part of a social-interactive process, not so much related to an individual's inner emotions but rather to the structure and content of a conversation (Bavelas, Gerwing, & Healing, 2014a).

More recently, more and more researchers have turned towards considering facial movements as *communicative* signals, for example in the context of depictions, where facial gestures can serve to "stage a scene" (Clark, 2016), for example to impersonate a particular character when telling a story (see also "reenactment", Sidnell, 2006; "facial portrayal", Bavelas, Gerwing, & Healing, 2014b; "multimodal quotation", Stec, Huiskes, & Redeker, 2015). And of course, gaze direction has long been acknowledged to play a fundamental role in signaling communicative intentions (e.g., Argyle & Cook, 1976; Senju & Johnson, 2009).

Second only to gaze shifts, some of the most prevalent facial movements in conversation are eyebrow movements such as eyebrow raises and furrows. According to the Facial Action Coding System (FACS: Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002)—an anatomically-based system that allows for coding of visually distinguishable facial movements (termed Action Units [AU])—eyebrow raises are realized by the Inner Brow Raiser (Central Frontalis; AU1) together with the Outer Brow Raiser (Lateral Frontalis; AU 2) while eyebrow furrows are realized by the Brow Lowerer (Corrugator, Depressor Supercilli, Depressor Glabellae; AU4; see Figure 1).

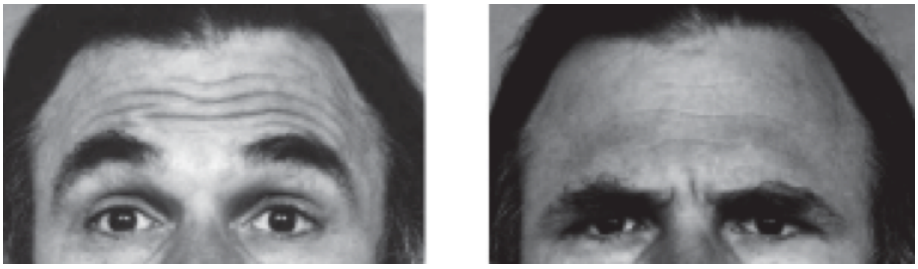

Figure 1. Example images of an eyebrow raise (AU 1+2) and an eyebrow furrow (AU 4) (Ekman, Friesen & Hager, 2002).

In the emotion domain, eyebrow raises have been associated with positive emotions like surprise, and eyebrow furrows with negative emotions like anger (Ekman, 1993). In terms of non-emotional signaling, eyebrow movements have been thought to occur in requests for information from a conversational partner (Darwin, 1872; Eibl-Eibesfeldt, 1972; Ekman & Friesen, 1975; Ekman, 1979; Wierzbicka, 1999). Indeed, eyebrow position is a grammaticalized facial question marker in many sign languages (Baker-Shenk, 1983; Coerts, 1992; Zeshan, 2004; Dachkovsky & Sandler, 2009). Specifically, eyebrow movements have been shown to fulfill an important conventionalized signaling function in signed languages in a particular type of question context which is core to achieving and sustaining mutual understanding in conversation—so-called 'other-initiated repair' (OIR) (Manrique, 2016). OIR is a brief exchange that momentarily interrupts the progress of a conversation to solve a communicative problem (Schegloff, Jefferson, & Sacks, 1977). An OIR sequence consist of a *repair initiation*, a signal from the recipient of a problem in perceiving or understanding what the speaker just said, and a *repair solution*, involving the speaker repeating part or all of the trouble source turn, clarifying certain parts of it, or confirming or disconfirming a candidate understanding offered by the recipient (Dingemanse, Kendrick, & Enfield, 2015).

Judging from their linguistic functions in sign languages, eyebrow raises and furrows may also be normative practices in spoken OIR. While repair can be initiated and resolved with spoken language in the absence of the visual channel (think of speaking on the phone), in spoken face-to-face conversation eyebrow raises and furrows have also been observed in OIR contexts (Enfield et al., 2013). An open question is whether eyebrow movements play a communicative role in initiating repair in spoken languages, or whether they might be epiphenomenal, that is, mere correlates or "ornaments" of verbal initiations without a signaling function in their own right.

A few studies provide initial clues that eyebrow movements may not be epiphenomenal in spoken OIR (Dingemanse, 2015; Kendrick, 2015; Floyd, Manrique, Rossi, & Torreira, 2016). First, comparing OIR sequences in unrelated spoken- and signed languages (Northern Italian, Cha'palaa, Argentine Sign Language), Floyd and colleagues (2016) showed that if a repair initiation was accompanied by a bodily "hold", that is, if body movements like eyebrow movements (but also, e.g., hand gestures or head movements) were "temporarily and meaningfully held static" (Floyd et al., 2016; p. 1), this hold was not disengaged until the communicative problem was solved. Floyd et al. (2016, p. 187) interpreted these holds as displaying that a repair solution is still expected, whereas disengaging from a hold displays that a repair solution is no longer expected because one has been provided. This indicates that disengaging from a brow position may play a communicative role in signaling successful grounding (i.e., through accomplishing closure of an OIR sequence), which may suggest that brow movements during repair *initiation* may be communicative too. Note that Floyd et al. (2016) did not distinguish between different types of brow movements such as furrows versus raises, though. Second, two individual descriptive examples—one from English ("raises her eyebrows, pulls down the corner of the mouth"; Kendrick, 2015, p. 11) and one from Siwu ("puzzled look: furrowing of eyebrows", Dingemanse, 2015, p. 238)—suggest that facial signals including eyebrow raises or furrows can be treated as repair initiations without relying on accompanying verbal material. While these studies reviewed above suggest that eyebrow movements may serve a communicative role in initiating repair both in signed as well as spoken language, little is known about the different compositions of repair initiations used in spoken language (e.g., verbal signal with versus without eyebrow movement) and about the functions of different types of eyebrow movements, such as brow raises and furrows.

Darwin (1872) proposed in his principle of antithesis that two opposed movements are likely to develop distinct communicative functions. Eyebrow raises and furrows are formally opposed, constituting two maximally contrastive extremes of how eyebrows can move. They have distinct effects on vision (Darwin, 1872; Lee et al., 2014), and, as mentioned above, they have been associated with emotions of opposed valence (Ekman, 1993). Assuming that eyebrow movements have a signaling function, this raises the question of whether eyebrow raises and furrows may also serve distinct communicative functions in signaling problems of perceiving or understanding in spoken face-to-face conversation. In Dutch Sign Language (Sign Language of the Netherlands [NGT], Coerts, 1992), eyebrow raises mark polar questions (e.g. 'you mean John?') and eyebrow furrows mark content questions (e.g., 'who?'). If the non-obligatory use of eyebrow actions in information requests in spoken Dutch is akin to the grammatically obligatory use of eyebrow actions in requests for information in Dutch Sign Language, one may expect that in spoken Dutch, eyebrow raises may be more often involved in repair initiations that make confirmation or disconfirmation relevant (e.g., 'you mean John?') and eyebrow furrows more often in repair initiations that make clarification relevant (often including content question words, e.g., '*Which* John?'). While the type of brow movement involved in a multimodal repair initiation may affect which type of repair solution is provided in response, the mere presence of the brow movement and the timing of the brow movement relative to the verbal signal in repair initiations may affect the speed by which a repair solution is provided. If the brow movement is initiated before the verbal signal, it may "forewarn" the speaker about a communicative problem, providing the speaker with a timing advantage when planning an appropriate response (see also Kaukomaa, Peräkylä, & Ruusuvuori, 2014, on how turn-opening frowns can anticipate utterances involving some kind of trouble, e.g., epistemic challenge).

In the present study, we hypothesize that eyebrow actions contribute to signaling problems of perceiving or understanding in spoken languages just as they do in sign languages, on the grounds that spoken languages also strongly rely on the visual channel, at least in face-to-face contexts (e.g., Clark, 1996; Bavelas & Chovil, 2000; Kendon, 2004; Enfield, 2009; Bavelas & Healing, 2013). We also hypothesize that eyebrow raises and furrows may serve different functions in signaling problems of perceiving or understanding. Specifically, we predict

(1) the type of eyebrow action used with verbal repair initiations to be associated with the type of repair solution provided in response (e.g., confirmation vs. clarification),

(2) repair time to be reduced by the presence of an eyebrow action or by an eyebrow action produced as a preliminary to verbal repair initiations, and

(3) addressee eyebrow actions alone, that is, silently produced in the absence of on-record verbal repair initiations, to also occasion repair.

To address these issues, we used two corpora of dyadic Dutch face-to-face conversations, which were specifically designed for detailed analyses of facial behavior [13]. We identified OIR sequences in conversations and coded the compositionality of repair initiations, focusing on eyebrow raises and furrows. We then quantified the co-occurrence of different linguistic formats of verbal repair initiations with eyebrow raises and furrows, the temporal relationship between the visual and the verbal component in the multimodal repair initiations, and investigated whether the presence of eyebrow actions in general and early eyebrow actions in particular (produced as preliminaries to verbal repair initiations) speed up the repair process. Finally, we identified silently produced addressee eyebrow raises and furrows that were treated as making relevant repair despite the absence of on-record verbal repair initiations.

Methods

Participants and corpora

We used two corpora of spontaneous, dyadic Dutch face-to-face conversations: the IFA Dialog Video Corpus (IFADV; van Son, Wesseling, Sanders & Heuvel, 2008) and the purpose built corpus of Dutch Face-to-Face (DF2F) conversation (see also Hömke, Holler & Levinson, 2017). Both corpora were specifically designed to allow for detailed analyses of communicative facial behavior.

The IFADV Corpus consists of 23 dyads, all native Dutch speakers (12-72 years) who knew each other well prior to the recording. Nine of the dyads consisted of a female and a male participant, 11 were all female, and three were all male. Five of the participants participated in two dyads each. The dyads were engaged in spontaneous Dutch face-to-face conversations for 15 minutes each. Conversations were recorded in a soundproof room and participants were seated at a table, facing each other, positioned approximately 1 m from each other (see Supplementary Material 1). Two

[13] One limitation of previous studies on eyebrow movements was that the corpora they used were not suitable for detailed analyses and quantification of facial signals. As Kendrick (2015) notes, “In some cases, the relevant participant is off-camera or his or her face cannot be seen due to the angle of the camera” (p.11). Floyd et al. (2016) pointed out that “speakers’ faces were not always clearly visible in the video” (p.190).

video cameras (JVC TK-C1480B, 720x576, 25 fps) were used to record frontal views of each participant and audio was recorded using head-mounted microphones (Samson QV).

The DF2F corpus consists of 10 dyads, all native Dutch speakers (18–68 years) who knew each other well prior to the recording. Four of the dyads consisted of a female and a male participant, four were all female, and two were all male. Each participant participated only in one dyad. The dyads were engaged in casual Dutch face-to-face conversations for 1 hour each and the recordings took place at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. The recordings took place in a soundproof room and participants were positioned approximately 1 m from each other at a 45-degree angle. Three HD video cameras (JVC GY-HM100) were used to record frontal views of each participant (see Supplementary Material 2) and a scene view. Audio was recorded using lightweight head-mounted microphones (DPA-d:fine-88) and an audio recorder (Roland R-44) recorded the two audio tracks in synchrony. Each recording session resulted in three videos and two audio files, which were then synchronized and exported in Adobe Premier Pro CS6 (MP4, 24 fps). Each recording session consisted of three 20-minute phases: During one 20-minute phase, participants did not wear the head-mounted microphones and audio was only recorded using a ceiling microphone. During a second 20-minute phase, audio was recorded using the head-mounted microphones, and during a third 20-minute phase, audio was recorded using the head-mounted microphones and, in addition, participants wore eye-tracking glasses. To achieve the highest audio quality and to allow for detailed analyses of facial behavior (without potential occlusion of or interference with facial behavior related to wearing eye-tracking glasses), only the 20-minute phase in which participants wore head-mounted microphones was used for this study. Each participant was paid 16 euros for the whole session which lasted about 90 minutes. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen, and informed consent was obtained before and after filming.

Analysis

We identified occurrences of other-initiated repair and eyebrow raises and furrows, sampling from randomly selected 10-minute segments in the IFADV corpus (one segment per dyad, resulting in 230 minutes) and from naturally occurring tellings

(Mandelbaum, 2013) in the DF2F corpus (all tellings in all dyads, resulting in 68 minutes), resulting in a total of 298 minutes of conversation. The choice to sample from randomly selected segments in the IFADV corpus and tellings in the DF2F corpus was based on practical considerations. OIR cases were already partially coded in the IFADV corpus (by PH) and brow movements were already partially coded in tellings of the DF2F corpus (by PH) [and we had no reason to assume systematic differences in the use of eyebrow raises and furrows between these two types of selected conversational materials].

Verbal other-initiated repair. We first focused the analysis on verbal cases of other-initiated repair, i.e., sequences "in which a turn T0 signals some trouble in a prior turn T-1 and is treated as making relevant the provision or ratification of a repair solution in a next turn T+1" (Dingemanse & Enfield, 2015, p. 99). For each OIR case, the linguistic format of the verbal repair initiation as well as non-mutually exclusive characteristics of the verbal repair solution was coded. Three basic formats of repair initiations were distinguished (Dingemanse & Enfield, 2015). A repair initiation was coded as (1) *open request* if it targeted the prior turn as a whole (e.g., *huh?*), typically making repetition relevant but sometimes also clarification, (2) as a *restricted request* if it targeted a specific aspect of the prior turn (e.g., *who?*), making clarification of this aspect relevant, and (3) as *restricted offer* if it targeted a specific aspect of the prior turn by offering a candidate understanding (e.g., *you mean John?*), making confirmation or disconfirmation relevant. For each repair solution, it was coded whether any material from the trouble source turn was (1) repeated, (2) clarified, or whether (3) it included a confirmation or disconfirmation (non-mutually exclusive options). A repair solution was coded as 'repeating' if some or all material from the trouble source turn was repeated (Curl, 2005), not taking into account whether 'dispensable' items such as a turn-initial *but* or *oh* (Schegloff, 2004) was omitted or not. A repair solution was coded as 'clarifying' if it involved modification or specification of the trouble source (Mazeland & Zaman-Zadeh, 2004), that is, if some or all material from the trouble source was rephrased, replaced, or if something was added. A repair solution was only coded as '(dis)confirming' if it included a "'yes/no/indeed' type item, a head nod/shake, or a repetition (+/- negation)" (Dingemanse, Kendrick, & Enfield, 2016; p. 42), often produced in response to an offered candidate understanding as part of the repair initiation (Schegloff, Jefferson,

& Sacks 1977; Schegloff, 2004). Note that a repair solution was not coded as '(dis)confirming' if it included an indirect (dis)confirmation, for example, by offering an alternative.

Criteria for identifying and classifying OIR cases were based on a coding scheme developed by Dingemanse et al. (2016). All repair sequences were identified by the first author (PH) experienced in the application of this coding scheme and resulted in a total of 116 OIR cases. Thus, there was a repair initiation about once every 2.5 minutes. This frequency is lower than the frequency of once every 1.6 minutes previously reported based on a large-scale cross-linguistic study of OIR (Dingemanse et al., 2015). While both studies focused on maximally informal conversations suggesting a similar amount of shared knowledge among participants, this difference in frequency may be due to the fact that participants in the study by Dingemanse et al. (2015) were often engaged in parallel activities such as preparing food, eating, or playing games, potentially leading to more problems in hearing or understanding due to background noises and distractions. In contrast, both corpora used for the current study were recorded in soundproof laboratories with little to no background noises or visual interference, let alone opportunities for potentially distracting parallel activities.

Eyebrow actions. We identified eyebrow raises and furrows (see Facial Action Units 1+2 and 4, respectively; Ekman & Friesen, 1978), annotated from the first to the last visible movement of the eyebrows. Eyebrow actions were identified by two independent coders (KK and MK) who were blind to the hypotheses. Twelve minutes were coded for training and 59 randomly selected minutes (approx. 20% of the total data) were coded for measuring inter-rater reliability. The inter-rater reliability was 76.5 % for brow action occurrence and a Cohen's Kappa (Landis & Koch, 1977; Holle & Rein, 2015) of .88 was achieved for agreement about the brow action type (brow furrow versus brow raise) indicating substantial agreement.

Compositionality of repair initiations. For each repair sequence, firstly, we assessed whether the verbal repair initiation co-occurred with eyebrow actions or not. Secondly, if verbal repair initiations co-occurred with eyebrow actions, we assessed the temporal relationship between the visual and the verbal component. Eyebrow actions were considered to be "co-occurring" if the eyebrow action temporally overlapped with a verbal repair initiation. Eyebrow actions were also considered to be “co-occurring” if the offset of the eyebrow action immediately preceded the onset of the verbal repair initiation without perceptible interruption or if the onset of the eyebrow action immediately followed the offset of the verbal repair initiation without perceptible interruption, such that the behaviors together formed a multimodal *Gestalt* (Mondada, 2014). More precisely, if the onset of the verbal repair initiation and the onset of the eyebrow action coincided precisely or if the onset of one preceded the onset of the other by less than 200 ms (up to which it is likely perceived as synchronous, as has been established for visible lip movements and articulatory sound; McGurk & MacDonald, 1976) this was coded as “initiated simultaneously”. If the onset of the eyebrow action preceded the onset of the verbal repair initiation by more than 200 ms this was coded as ‘initiated visually first’ (or ‘verbal OIR with visual preliminary’, see Results section below), and if the onset of the verbal repair initiation preceded the onset of the eyebrow action by more than 200 ms, it was coded as ‘initiated verbally first’.

Eyebrow actions occasioning repair in the absence of vocalization. Finally, when eyebrow actions alone were sufficient to occasion repair, that is, without any ‘on-record’ verbal repair initiation (e.g., Kendrick, 2015), they were coded as ‘eyebrow actions only occasioning repair’.

All annotations were created in ELAN 4.8.1 (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006).

Results

We focus the analysis first on the compositionality of repair initiations, assessing the co-occurrence of different eyebrow actions with different linguistic formats of the verbal repair initiation. We then explore the corpus-based plausibility of whether eyebrow actions might merely be epiphenomena of verbal repair initiations or whether they may contribute to signaling problems in hearing or understanding by examining (1) whether the type of eyebrow action accompanying repair initiation predicts certain types of repair solutions, even after taking into account variability in the co-occurring verbal repair initiation format, (2) whether the presence of an eyebrow action as a preliminary to repair initiation speeds up the repair process, and finally, (3) whether eyebrow actions alone are sufficient to occasion repair.

Initiating repair with words and brows

Out of all identified verbal repair initiations (N=116), a substantial number co-occurred with eyebrow actions (40% [n=46]). Out of those co-occurring with eyebrow actions, about half co-occurred with eyebrow raises (46% [n=22]) and the other half with eyebrow furrows (54% [n=25]).

Which composition (verbal-only repair initiation, verbal repair initiation with eyebrow raise, or verbal repair initiation with eyebrow furrow) co-occurred with which linguistic format of the verbal repair initiation (open request, restricted request, restricted offer)? As one can see in Figure 1, restricted offer was the overall most frequent linguistic format of the verbal repair initiation (68% [n=72]), followed by restricted request (24% [n=25]) and open request (8% [n=9]). While the distribution of linguistic formats of the verbal repair initiation is almost identical when considering just verbal-only repair initiations (restricted offer: 72% [n=44]; restricted request: 18% [n=11]; open request: 10% [n=6]) and just verbal repair initiations with eyebrow raises (restricted offer: 71% [n=15]; restricted request: 19% [n=4]; open request: 10% [n=2]), verbal repair initiations with eyebrow furrows show a lower proportion of restricted offers (54% [n=13]) and open requests (4% [n=1]), but a substantially higher proportion of restricted requests (42% [n=10]), relative to verbal-only repair initiations and verbal repair initiations with eyebrow raises[14].

[14] Note that two rare linguistic formats of restricted OIR were excluded: alternative questions (invites a selection from among alternatives; n(with raise)=0, n(with furrow)=0, n(verbal without eyebrow action)=2) and external repair initiations (address problems about unexpressed elements of T-1; n(with raise)=1, n(with furrow)=1, n(verbal without eyebrow action)=6), resulting in a total of 106 OIR cases.

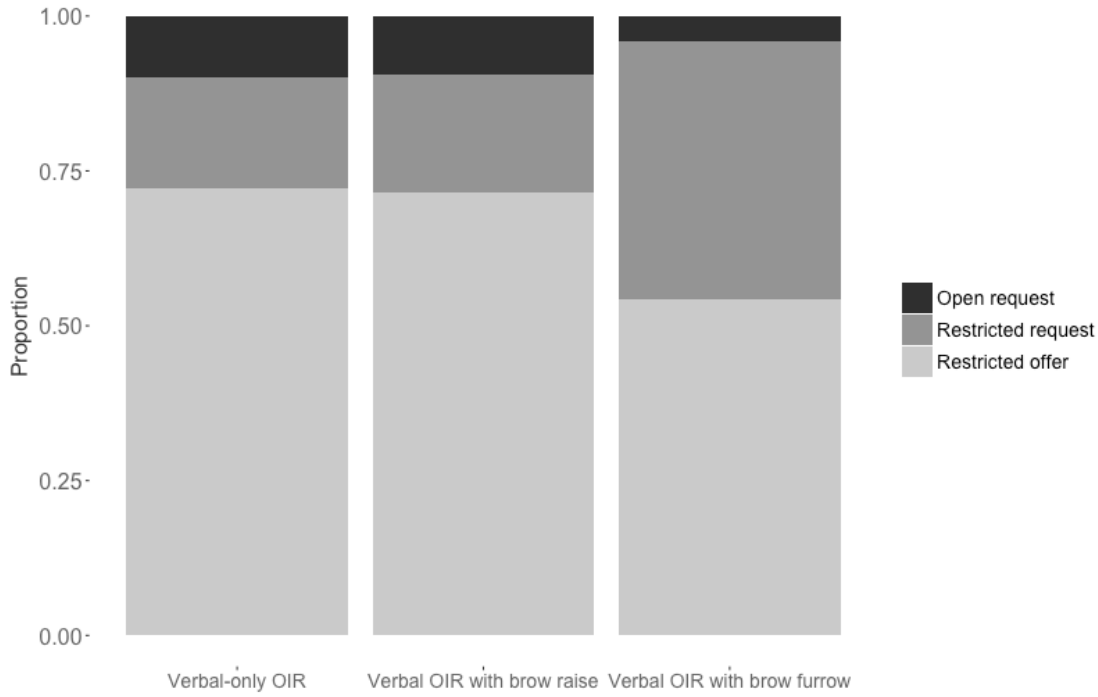

Figure 1. Compositionality of repair initiations (verbal-only OIR, verbal OIR with eyebrow raise, verbal OIR with eyebrow furrow) by linguistic format of the verbal repair initiation (open request, restricted request, restricted offer; N=106).

To test whether the *presence* of eyebrow action (verbal-only repair initiation, verbal repair initiation with eyebrow action) in repair initiations is statistically associated with the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer), a mixed effects logistic regression analysis was performed (including random intercepts for participants). An intercept-only model was compared to a model in which ‘presence of eyebrow action’ was added as a predictor variable, using a Likelihood Ratio Test. Including ‘presence of eyebrow action’ did not improve the model fit significantly ($\chi^2(2) = 2.56$, $p = .276$), indicating that the presence of eyebrow action did not reliably distinguish between the linguistic format of the verbal repair initiation. To test whether the *type* of eyebrow action (verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) in repair initiations is associated with the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer), an additional mixed effects logistic regression analysis was performed (including random intercepts for participants). An intercept-only model was compared to a model in which ‘type of eyebrow action’ was added as a predictor variable, using a Likelihood Ratio Test. Including ‘type of eyebrow action’ did not improve the model fit significantly ($\chi^2(2) = 1.88$, $p = .389$), indicating that the type of eyebrow action did not reliably distinguish between the linguistic format of the verbal repair initiation.

The results reveal that a substantial number of verbal repair initiations are accompanied by eyebrow actions—about as often by eyebrow raises as by eyebrow furrows. Furthermore, the results numerically mirror the hypothesized pattern based on question marking in Dutch sign language that eyebrow raises may be more often involved in repair initiations that make confirmation or disconfirmation relevant (restricted offers like ‘You mean John Smith?’) and eyebrow furrows more often in repair initiations that make clarification relevant (like restricted requests such as ‘Who?’, Coerts, 1992): relative to eyebrow furrows, a larger proportion of eyebrow raises accompanied restricted offers, and relative to eyebrow raises, a larger proportion of eyebrow furrows accompanied restricted requests. However, these differences were not statistically significant.

The example below illustrates how an eyebrow raise can be used with a restricted offer, which is subsequently confirmed through a head nod:

```
DF2F corpus_19_266591

1 A:  Hij heeft nu de vriendin van Boris
      he has now the girlfriend of Boris
      He now has Boris' girlfriend

2 B:  Ja m- (.) ((raises brows, see Figure 2)) Jeanette?
      yeah m-                                  Jeanette?
      yeah m-                                  Jeanette?

3 A:  ((nods))
```

Figure 2. Eyebrow raise produced with a restricted offer as linguistic format of the verbal repair initiation (‘Jeanette?’, line 2 in the example above; see video in Supplemenatary Material 3)

By contrast, the following example illustrates how an eyebrow furrow can be used with a restricted request, in this case for clarification of an underspecified person reference, which is subsequently provided[15].

```
IFADV_17_588780

1 A: Ik ben dus achternichtje met Marieke
     I  am  thus second cousin with Marieke
     It turns out I’m second cousin of Marieke

2 B: ((furrows brows, see Figure 2)) Marieke, wie is Marieke?
                                     Marieke, who is Marieke?
                                     Marieke, who is Marieke?

3 A: Ja, die ene van de Kleinkunst
     Yeah, the one from the cabaret
     Yeah, the one from cabaret
```

[15] See Dingemanse, 2015, Extract 5, for a strikingly similar example including eyebrow furrowing in Siwu, an African language spoken in a small community in eastern Ghana.

Figure 3. Eyebrow furrow produced with a restricted request as linguistic format of the verbal repair initiation ('who is Marieke?', line 2 in the example above).

Do eyebrow actions contribute to initiating repair in spoken face-to-face conversation?

On the one hand, the co-occurrence of eyebrow actions and verbal repair initiations suggests they may serve similar functions, that is, they may be co-expressive in signaling problems in hearing or understanding. On the other hand, it raises the question whether the eyebrow actions in these cases might be epiphenomenal, and thus merely correlates but not functionally involved in signaling problems in hearing or understanding. Below, we present three pieces of evidence suggesting that eyebrow actions are indeed effective in signaling problems in hearing or understanding: First, we show that verbal repair initiations with eyebrow furrows are more likely to get clarifications as repair solutions compared to either verbal repair initiations with eyebrow raises or verbal-only repair initiations, even after taking into account variance explained by the linguistic format of the co-occurring verbal repair initiation, pointing to a potentially unique contribution of eyebrow furrows in signaling a need for clarification (relative to either verbal repair initiations with eyebrow raises or verbal-only repair initiations). Secondly, we show that, relative to repair initiations without eyebrow actions, repair initiations that were immediately preceded by eyebrow actions as preliminaries get repaired faster. Finally, and most importantly, we show that eyebrow furrows alone can be sufficient to occasion clarification. We take up these three lines of evidence in order:

(1) Does the presence or type of eyebrow action in repair initiation predict the type of solution provided? To address this question, we correlated the composition of the repair initiation (verbal-only repair initiation, verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) with different non-mutually exclusive characteristics of the subsequent repair solution, namely whether any material from the trouble source turn was repeated, clarified, or whether it included a confirmation or disconfirmation. Note that this analysis could not be applied to six OIR cases in which the T+1 was absent (N=100)[16]. Repair solutions in response to verbal-only repair initiations were slightly more likely to include repetitions (38% [n=25]) than repair solutions in response to verbal repair initiations with eyebrow raises (33% [n=7]), and slightly less likely than repair solutions in response to verbal repair initiations with eyebrow furrows (42% [n=10]). Repair solutions in response to verbal-only repair initiations were more likely to include (dis)confirmation (71% [n=46]) than repair solutions in response to verbal repair initiations with eyebrow raises (57% [n=12]), and slightly less likely than repair solutions in response to verbal repair initiations with eyebrow furrows (50% [n=12]). As one can see in Figure 4, repair solutions in response to repair initiations with eyebrow furrows were more than twice as likely to include clarification (65% [n=15]) relative to repair solutions in response to verbal-only repair initiations (31% [n=18]) and repair initiations with eyebrow raises (25% [n=5]).

[16] Out of the six OIR cases in which the T+1 was absent, four were verbal-only repair initiations, one with a brow furrow, and one with a brow raise. That is, multimodal repair initiations were ignored less often, potentially suggesting that repair initiations are “weaker” if they are not produced with eyebrow actions (but bear in mind the small number of cases).

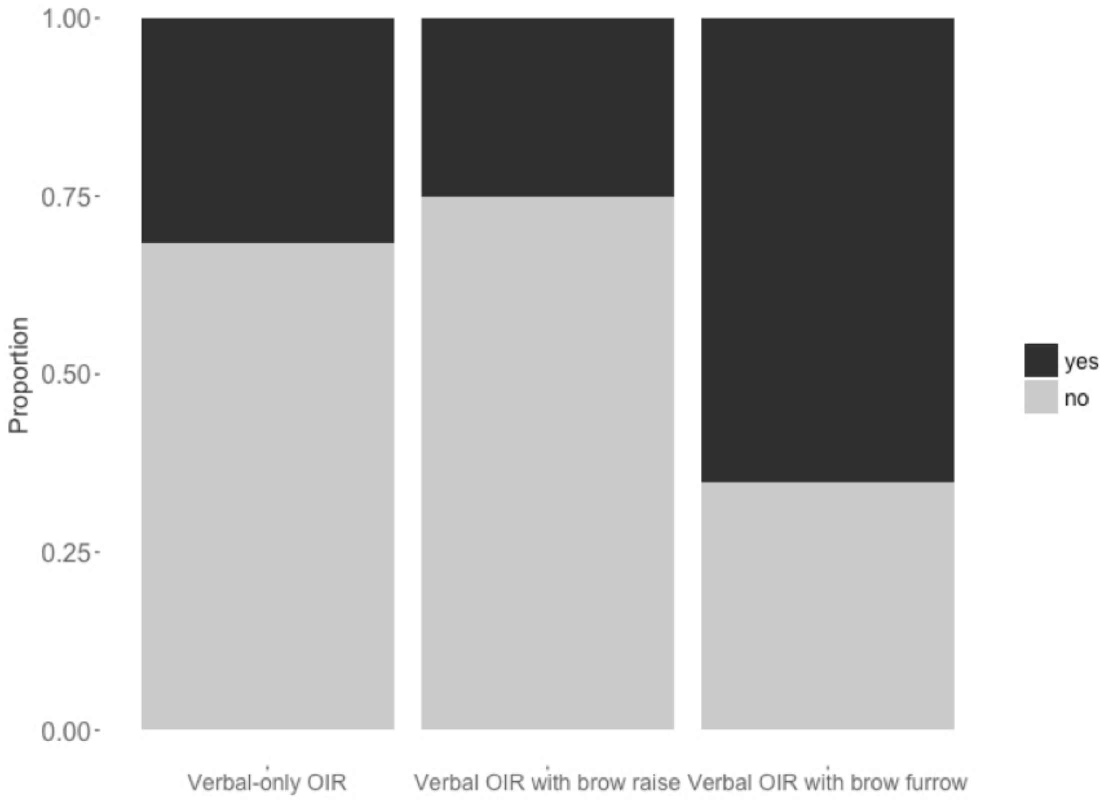

Figure 4. Compositionality of OIR (verbal-only OIR, verbal OIR with eyebrow raise, verbal OIR with eyebrow furrow) by repair solution (YES = with clarification, NO = without clarification), N=100.

Note, however, that this is not necessarily a unique contribution of eyebrow furrows. One might argue that given that eyebrow furrows are more frequent in restricted requests (see Figure 1), it is not surprising that repair solutions in response to repair initiations with furrows are more likely to include clarification. The linguistic format of the verbal repair initiation, in this case 'restricted request', rather than the accompanying eyebrow furrow, may thus underlie the increased likelihood for repair solutions to include clarifications. To explore this possibility, we used a mixed effects logistic regression analysis (including random intercepts for participants) to test whether the composition of the repair initiation (verbal-only repair initiation, verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) predicts whether the repair solution included a clarification or not (clarification, no clarification), while taking into account variability in the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer; see Figure 1) by adding it as a predictor variable to the statistical model. This model was compared to a reduced model without the predictor variable of ‘composition of repair initiation’ using a Likelihood Ratio Test. Including ‘composition of repair initiation’ improved the model fit significantly ($\chi^2(2) = 7.85$, $p<.05$), revealing that, relative to repair initiations with eyebrow raises, repair initiations with eyebrow furrow changes the log odds of a subsequent repair solution including clarification by 1.66 ± 0.71 (standard error), and relative to repair initiations without eyebrow actions, repair initiations with eyebrow furrow changes the log odds of a subsequent repair solution including clarification by 1.38 ± 0.55 (standard error). These results indicate that, independently of the linguistic format of the repair initiation, the presence of an eyebrow furrow increased the likelihood of a repair initiation to be treated as a request for clarification.

(2) Do eyebrow actions speed up the repair process? *Presence of eyebrow action: Verbal repair initiation with versus without eyebrow action.* If eyebrow actions were merely a correlate of verbal repair initiation—say a *symptom* of cognitive effort—rather than a communicative *signal* of a problem in hearing or understanding, one should expect the repair time, measured from the end of the repair initiation to the start of the repair solution, to be unaffected by whether the repair initiation was produced with or without an eyebrow action. Alternatively, if eyebrow actions can indeed function as a communicative *signal* of a problem in hearing or understanding, one may expect that—by increasing redundancy—the presence of an eyebrow action per se may reduce potential ambiguity and express a stronger sense of urgency, which may reduce the repair time. To address this issue, we compared the repair time between verbal repair initiations without versus with a brow action (see Figure 5).

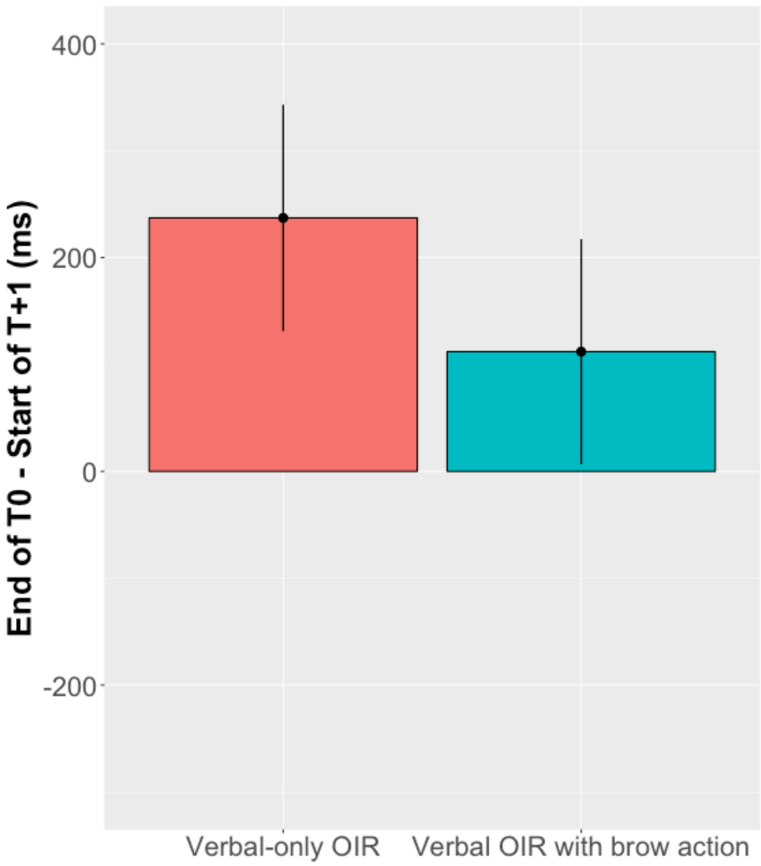

Figure 5. Repair time, measured from the end of the repair initiation (T0) until the start of the repair solution (T+1), by repair initiation *without* an eyebrow action ('Verbal-only OIR') versus *with* an eyebrow action ('Verbal OIR with brow action'). Standard errors are represented in the figure by the error bars attached to each column.

We used R and *lme4* (Bates, Maechler, & Bolker, 2012) to test in a mixed-effects model whether repair time differed between verbal repair initiations with versus without a brow action, while taking into account variability in the linguistic format of the verbal repair initiation by adding it as a predictor variable to the statistical model. We entered 'linguistic format' (open request, restricted request, restricted offer) and 'presence of brow action' (yes, no) as fixed effects and intercepts for participants as a random effect into the model. This model was compared to a reduced model without 'presence of brow action' as a fixed effect using a Likelihood Ratio Test. Including 'presence of brow action' as a fixed effect improved the model fit marginally ($\chi^2(1)$ = 336, p = .066), revealing that—relative to repair initiations *without* a brow action (376.11 ms [mean] ± 142.48 [standard error])—the repair time for repair initiations *with* a brow action was shorter by about 135.73 ms (mean) ± 75.13 (standard error), but not reliably so (t = -1.806, p = .074). Note that adding ‘brow action type’ (brow raise, brow furrow) as a predictor to the statistical model did not improve the fit ($\chi^2(1) = 0.60$, p = .436), indicating that repair time was unaffected by the type of brow action involved.

Timing of eyebrow action: Verbal-only repair initiation versus repair initiation with concurrent versus early eyebrow action. If eyebrow actions were merely a correlate of verbal repair initiation—say a *symptom* of cognitive effort—rather than a communicative *signal* of a problem in hearing or understanding, one should expect the repair time, measured from the end of the repair initiation to the start of the repair solution, to be unaffected by whether the repair initiation was a verbal-only repair initiation or whether it was produced with a concurrent versus an early eyebrow action (i.e., produced as a visual preliminary initiated immediately before the verbal repair initiation). Alternatively, if eyebrow actions can indeed function as a communicative *signal* of a problem in hearing or understanding, one may expect that an early eyebrow action produced as a visual preliminary, a potential visual “forewarning”, may facilitate a timely response, thus reducing the repair time. To address this issue, we examined the temporal relationship between the visual and the verbal component in multimodal repair initiations (see Methods section) and then compared the repair time between *verbal-only* repair initiations, verbal repair initiations with a *concurrent* eyebrow action versus verbal repair initiations with an *early* eyebrow action (produced as a visual preliminary to the verbal repair initiation; see Figure 6).

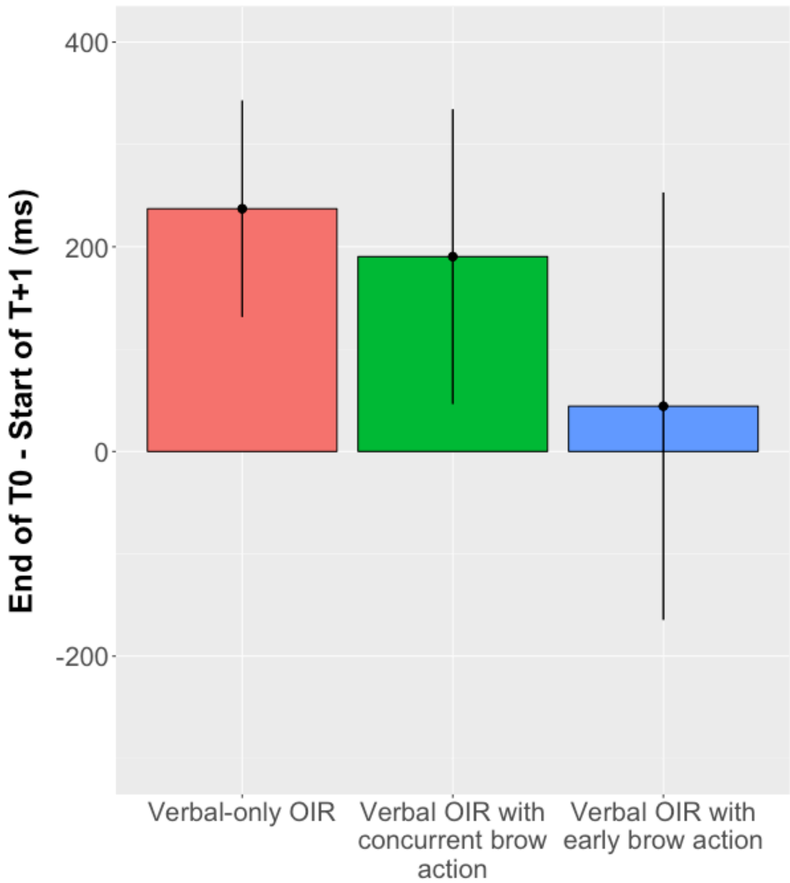

Figure 6. Repair time, measured from the end of the repair initiation (T0) until the start of the repair solution (T+1), by *verbal-only* repair initiations ('Verbal-only OIR') versus repair initiations with a *concurrent* eyebrow action ('Verbal OIR with concurrent brow action') versus an *early* eyebrow action (produced as a visual preliminary to the verbal repair initiation, 'Verbal OIR with early brow action'). Standard errors are represented in the figure by the error bars attached to each column.

We tested in a mixed-effects model whether repair time differed between verbal-only repair initiations versus repair initiations with a concurrent eyebrow versus an early eyebrow action, while taking into account variability in the linguistic format of the verbal repair initiation by adding it as a predictor variable to the statistical model. We entered 'linguistic format' (open request, restricted request, restricted offer) and 'brow action' (verbal-only repair initiation, repair initiation with early brow action, repair initiation with concurrent brow action) as fixed effects and intercepts for participants as a random effect into the model. This model was compared to a reduced model without 'brow action' as a fixed effect using a Likelihood Ratio Test.

Including 'brow action' as a fixed effect improved the model fit marginally ($\chi^2(3)$ = 6.49, p = .089). This revealed that the repair time for verbal repair initiations with a *concurrent* brow action was shorter by about 42.08 ms (mean) ± 92.95 (standard error), compared to *verbal-only* repair initiations (413.05 ms [mean] ± 141.79 [standard error]), and that this difference was not reliable (t = -0.453, p = .651). Also, the repair time for verbal repair initiations with an *early* brow action was shorter by about 189.58 ms (mean) ± 124.45 (standard error), compared to verbal repair initiations with a *concurrent* brow action (370.97 ms [mean] ± 165.56 [standard error])—but, again, not reliably so (t = -1.523, p = .131). However, relative to *verbal-only* repair initiations, the repair time for verbal repair initiations with an *early* brow action was significantly reduced by about 231.66 ms (mean) ± 107.32 (standard error), t = -2.159, p = .033. Note that adding 'brow action type' (brow raise, brow furrow) as a predictor to the statistical model did not improve the fit ($\chi^2(1)$ = 0.54, p = .462).

We have seen that, while the mere presence of a brow action did not reliably speed up the repair process, the presence of an early brow action produced as a visual preliminary reduced the repair time significantly, compared to verbal-only repair initiations, again suggesting that eyebrow actions are effective in signaling problems of hearing or understanding.

(3) Can eyebrow actions alone signal problems in hearing or understanding? To address this question, we identified all silently produced eyebrow actions that occasioned repair. This resulted in eleven identified eyebrow furrows and zero eyebrow raises. None of these eyebrow furrows were treated as making (dis)confirmation relevant but *all* of them were treated as making clarification relevant (while three of these were also treated as making partial repetition relevant). Despite these observations resting on a small number of cases, the result quite convincingly suggests that eyebrow furrows alone can be sufficient as signaling a need for clarification, even in the absence of verbal repair initiations. The example below illustrates how an eyebrow furrow alone can occasion repair, as if it was a restricted verbal request for clarification (see video in Supplementary Material 4).

ETC13_151369

```
1 B: Ik heb het in mijn telefoon staan
     I have it  in my telephone stand
     I have it on my phone

2    in een vroeger bericht [van Floortje
     in an  earlier message from Floortje
     in an  earlier message from Floortje

3 A:                         [((furrows brows, see Figure 6))

4 B: hoe ze heet
     how she called
     what her name is

5 A: Ja ((unfurrows brows))
     yes
     yes
```

Figure 6. Eyebrow furrow alone occasioning clarification (‘what her name is’, line 4 in the example above; see video in Supplemenatary Material 4)

In the example above, B targets *it* (‘het’, line 1) as the trouble source by clarifying what *it* referred to through a repair *what her name is* (‘hoe ze heet’, line 4). As such, without any on-record verbal prompting, A’s eyebrow furrow was treated as if A had produced a verbal restricted request like *What do you have on your phone?* (“Wat heb je in je telefoon staan?”).

Discussion

Do eyebrow movements serve a communicative function in signaling problems of hearing or understanding in spoken conversation? The present findings suggest they do indeed. The results are incompatible with an epiphenomenal interpretation of eyebrow movements, because (1) in addition to the linguistic format of the verbal repair initiation, the type of co-occurring eyebrow movement independently predicted the type of repair solution, (2) the presence of an eyebrow movement as a visual preliminary to verbal repair initiations enhanced repair speed, and (3) eyebrow movements alone were sufficient to occasion clarification.

First, we have seen that eyebrow raises and furrows were both used with all three basic linguistic formats of repair initiation, whether the co-occurring repair initiation targeted the prior turn as a whole (open request), a specific aspect of it (restricted request), or whether the repair initiation offered a candidate understanding (restricted offer). A higher proportion of eyebrow furrows co-occurred with restricted requests (‘Which John?’) relative to repair initiations with eyebrow raises, and a higher proportion of eyebrow raises co-occurred with restricted offers (‘John Smith?’) relative to repair initiations with eyebrow furrows—a numerical pattern that parallels the linguistic function of eyebrow position in Dutch Sign Language, where eyebrow furrows serve as non-manual grammatical markers of content questions and eyebrow raises as non-manual grammatical markers of polar questions (Coerts, 1992). Bear in mind, however, that these numerical differences were not statistically significant. Repair initiations without eyebrow actions and repair initiations with eyebrow raises showed an almost identical distribution regarding restricted offers, potentially pointing to a higher optionality of the use of eyebrow raises in polar questions as repair initiations in spoken face-to-face conversation.

Second, we have also seen that the type of eyebrow movement co-occurring with repair initiations predicted differences in how these multimodal signals of problems were treated as making relevant different solutions. The presence of an eyebrow furrow uniquely—i.e., independently of the linguistic format of the verbal repair initiation—increased the likelihood of a repair initiation to get a repair solution including clarification. This suggests that the visual component is not merely a correlate of the verbal component, but that visual and the verbal can be co-expressive in multimodal repair initiations. More generally, while eyebrow movements are not necessary for initiating repair in spoken conversation, this result suggests that they can nevertheless serve a communicative function, contributing to signaling the type of communicative problem and how it can best be fixed.

Third, we have seen that the presence of eyebrow movements as visual preliminaries to repair initiations reduced repair time, relative to repair initiations without eyebrow movements as visual preliminaries (i.e., verbal-only repair initiations)[17]. The eyebrow movement as a preliminary to a repair initiation here seems to serve a similar signaling function for a speaker as the orange light as a preliminary to the red light in traffic lights for a driver. While the speaker can speak through the addressee's eyebrow movement as well as the driver can drive through the orange light—both without being sanctioned—these preliminaries seem to facilitate a timely response, serving as 'forewarnings' of an upcoming disruption of progress. As such, this result is in line with findings from other domains of human joint action in which 'making oneself predictable' facilitates coordination (e.g., Vesper, van der Wel, Knoblich, & Sebanz, 2011). More generally, it again suggests that eyebrow movements are effective in the context of initiating repair and it illustrates how visual signals may enhance communicative efficiency in spoken languages (see also Holler, Kendrick, & Levinson, 2017, on questions getting faster responses if accompanied by gesture).

Fourth, we have seen that eyebrow furrows alone can silently signal insufficient understanding. This result suggests that while off-record facial action like the eyebrow furrow is usually not considered to be part of turn-constructional units in the turn-taking system (Sacks, Schegloff, & Jefferson, 1974; Dingemanse & Floyd, 2014), it can serve sequentially equivalent functions as verbal repair initiations. As Levinson (2013) noted, "Words and deeds are the same kind of interactional currency" (p. 74). The eyebrow furrow could be considered an implicit or off-record type of

[17] Note that some of these results on repair time were based on a relatively small sample size (e.g., n[Verbal OIR with early brow action]=18). Further research validating its generalizability would be desirable.

other-initiation of repair. While an eyebrow furrow seems slightly more accountable than a ‘freeze look’ (Manrique, 2016), it still does not explicitly encode the intention to initiate repair—potentially in an effort to minimize any possible “face-threatening” consequences (Brown & Levinson, 1987)—“just as “It’s cold in here” does not explicitly encode the intention to get somebody to shut the window” (Manrique & Enfield, 2015, p. 11).

Moreover, purely visible bodily behaviors used to initiate repair have previously been classified as open requests (equivalent to e.g., *huh?*) as they do not explicitly target specific aspects of the trouble source but the trouble source as a whole, which is typically treated as making repetition relevant (e.g., Dingemanse et al., 2015). Accordingly, if eyebrow furrows served as open requests, one would have expected them to be typically treated as making repetition relevant. However, eyebrow furrows were treated specifically as making clarification relevant—even when not combined with a verbal repair initiation—suggesting that they may implicitly target certain aspects of the prior turn as in need of clarification. How could eyebrow furrows possibly target specific aspects of a prior turn to be clarified? On the one hand, if the eyebrow furrow as such signals a need for clarification, it may be easy to guess for the speaker, based on estimates of shared knowledge, which aspect of the prior turn needs clarification (e.g., an underspecified person reference). On the other hand, as with visual addressee signals more generally, since they do not interfere as much with the spoken turn as verbal addressee signals, a specific troublesome aspect of a turn cannot only be targeted through explicit verbal means (e.g., *Who?*) but also through timing. That is, producing the visual signal immediately after the troublesome part (e.g., ambiguous person reference) of the ongoing turn may already signal what part of the trouble source turn needs clarification. In-depth future examinations of the precise temporal relationship between brow movement and trouble source may shed light on this issue.

Note that in the present study we did not find any eyebrow *raises* that occasioned clarification without relying on a verbal signal, which may in part be explained by the close association of eyebrow raising and speaking (Krahmer & Sweerts, 2004; Flecha-García, 2010). While eyebrow furrows may intrinsically signal some kind of communicative trouble or puzzlement, eyebrow raises might be associated with verbal repair initiations, at least to some extent, because verbal repair initiations often have questioning prosody (e.g. *Huh?*; Enfield et al., 2013) and eyebrow raises can coexpress questioning prosody (Bolinger, 1983). This does not mean that eyebrow raises can never occasion repair without relying on a verbal signal in spoken Dutch. At least anecdotal evidence suggests that also in spoken Dutch, eyebrow raises—especially when combined with a downward movement of the corners of the mouth—can also occasion clarification without a verbal signal, especially after a try-marked person or place references. This facial gesture combining brow and mouth actions has been termed a "facial shrug" (Ekman, 1985; Bavelas et al., 2014)—a signal of "not knowing" (Bavelas et al., 2014, p. 15; see also Kendrick, 2015, p.10-11, for an example in English).

Taken together, the present results already provide suggestive correlational evidence. However, experimental work (Chapter 5) is required to provide conclusive evidence regarding the hypothesized causal involvement of eyebrow movements in signaling problems of hearing or understanding in spoken face-to-face communication.

We are suggesting that eyebrow movements serve a communicative function, but this does not necessarily entail that they are communicatively intended (Brennan, Galati, & Kuhlen, 2010). In fact, an eyebrow furrow might merely be a symptom of the addressees' processing difficulty or high cognitive load, which is then interpreted and treated by the speaker as indicating a need for clarification. Darwin (1872) already mentioned that eyebrow furrows (or 'frowns', as he called them) are not only associated with unpleasantness but also with a potentially related but distinct state of dealing with difficulty in thought:

"A man may be absorbed in the deepest thought, and his brow will remain smooth until he encounters some obstacle in his train of reasoning, or is interrupted by some disturbance, and then a frown passes like a shadow of his brow." (p. 221)

The observation that people—as individuals not engaged in conversation—also furrow their brows when dealing with cognitive difficulties suggests that such furrows may not only serve an other-oriented, communicative function in signaling a need for clarification in conversation, but that they may also serve a self-oriented, cognitive function (see Figure 7, for an illustration).

Photo on the left: retrieved from https://pxhere.com/en/photo/1127793, CCO.
Photo on the right: retrieved from https://www.flickr.com/photos/renaud-camus/8375622029, CCO, and cropped afterwards.

Figure 7. Rodin's sculpture *Le Penseur* ('The Thinker', 1880) and a facial close-up showing his furrowed eyebrows. Note that the philosopher Gilbert Ryle famously used *Le Penseur* in the mind-body debate, asking 'What is he doing?' (1968), arguing against the privacy of cognitive states.

Social-communicative functions and potential cognitive, perceptual, and emotional functions of eyebrow movements are not mutually exclusive. It is possible that the cognitive, perceptual, and emotional functions underlie and precede the communicative signaling function, phylogenetically as well as ontogenetically (e.g., Oster, 1978, reports eyebrow furrows during "concentration" already in one to three month old infants). The eyebrow furrow as a potential symptom of mental effort, for example, may have been co-opted for communicative purposes through processes of ritualization (Darwin, 1872; Tinbergen, 1952; Eibl-Eibesfeldt, 1972; Bruner, 1978; Tomasello, 2007), which would point to a non-arbitrary, iconic relationship (Grice, 1957; Perniss & Vigliocco, 2014) between form and function in communicative eyebrow furrows. In the same way in which closing the eyes by blinking may signal "no need to see anymore" because sufficient understanding has been reached (Hömke, Holler, & Levinson, 2017 / submitted), furrowing the eyebrows—as if trying to see more clearly[18,19]—appears to signal insufficient understanding, potentially shedding

[18] According to Darwin (1872), Prof. Donders already suggested that eyebrows are furrowed to see more clearly ("the corrugators are brought into action in causing the eyeball to advance in accommodation for proximity in vision", p. 221).

new light on the suggested "embodied" origin of the Understanding-Is-Seeing metaphor (Lakoff & Johnson, 1999) and on visual origins of mental-state signaling (Lee et al., 2014).

The results suggesting a communicative function of eyebrow movements in signaling informational needs in spoken Dutch are in line with examples from other spoken languages like English (Kendrick, 2015), Italian and Chapalaa (Floyd et al., 2014), and Siwu (Dingemanse, 2015), but also with studies on eyebrow movements in signed languages like Dutch Sign Language (Coerts, 1992; De Vos, Van Der Kooi, & Crasborn, 2009) and Argentine Sign Language (Floyd et al., 2014; Manrique, 2016). This suggests that eyebrow movements as signals of insufficient hearing or understanding may be independent from language modality—since they are used in spoken as well as signed language—as well as from language history—since they have been described in unrelated languages. If the use of eyebrow movements as a signal of insufficient hearing or understanding is stable across a variety of unrelated languages, this would be consistent with Darwin (1872) who noted "the Australians, Malays, Hindoos, and Kafirs of South Africa frown, when they are puzzled" and who suggested that "men of all races frown when they are in any way perplexed in thought" (p. 221), but it may also suggest that eyebrow movements as signals of communicative problems have evolved from common pressures of a shared conversational infrastructure (Levinson, 2006; Schegloff, 2006; Dingemanse, Torreira, & Enfield, 2013; Stivers et al., 2009; Levinson, 2016).

Acknowledgements

We thank Mark Dingemanse, Herb Clark, and Emma Valtersson for valuable discussions and Kim Koopmans (KK) and Marina Koleva (MK) for coding support. This research was funded by the European Research Council (Advanced Grant #269484 INTERACT awarded to Prof. Levinson) and the Max Planck Gesellschaft.

[19] See also Goodwin and Goodwin (1986) on the "thinking face", referring to the speaker marking a word search by turning away her gaze from the addressee with a distant look and with a facial gesture of someone thinking hard.

Chapter 5. The cooperative eyebrow furrow: A facial signal of insufficient understanding in face-to-face interaction

Unlike other animals, humans tend to face each other in everyday communication. This allows humans to rely not only on vocal but also on various visual bodily behaviors when communicating (Levinson & Holler, 2014). While the language sciences have made substantial progress in the study of hand gestures (e.g., McNeill, 2000; Kendon, 2004), there is one part of the body that has received relatively little attention despite its omnipresence in and intuitive relevance for everyday face-to-face communication: the face.

There is a large literature on *facial expressions* (Darwin, 1872; Ekman, 1993), which are often described as rather involuntary public manifestations of an individual's emotion, for example of fear upon seeing a spider. Facial expressions have been contrasted with more voluntary *facial gestures* (Bolinger, 1946; Kendon, 2004). Facial gestures are facial actions that are used as communicative signals, that are shaped by the structure and content of a social interaction rather than by an individual's emotional response (Bavelas, Gerwing, & Healing, 2014a; see also "conversational facial signals", Ekman, 1979, and "facial displays", Kraut & Johnston, 1979). Facial gestures have been shown to serve depictions (Clark, 2016), for example to impersonate a character when telling a story (see also "reenactment", Sidnell, 2006; "facial portrayal", Bavelas, Gerwing, & Healing, 2014b; "multimodal quotation", Stec, Huiskes, & Redeker, 2015). In addition, gaze direction serves a crucial role in signaling communicative intentions (e.g., Argyle & Cook, 1976; Senju & Johnson, 2009). Next to gaze shifts, some of the most frequent facial movements are eyebrow movements. In the emotion literature, eyebrow movements have been associated with distinct emotional states. While eyebrow raises have been shown to be linked to positive emotions (e.g., greetings, surprise), eyebrow furrows have been linked to negative emotions (e.g., anger; Ekman, 1993; see also Chapter 4, Figure 1, for example stills). At the same time, eyebrow movements have been proposed to be used as communicative signals, in requesting information from a conversational partner (Darwin, 1872; Eibl-Eibesfeldt, 1972; Ekman, 1979; see also Holler & Wilkin, 2011).

Our previous research has provided correlational corpus-based evidence suggesting that listener eyebrow furrows can indeed serve an interactional function in face-to-face communication (Chapter 4; Hömke, Holler, & Levinson, in prep), specifically to signal non-understanding. Signaling understanding and non-understanding is fundamental to "grounding" (Clark & Brennan, 1991), the process of establishing the mutual belief that communicative acts have been understood well enough for current purposes (Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989), which is crucial for successful communication. Hömke and colleagues (in prep) showed that verbal listener signals of non-understanding intended to elicit repair (such as *Huh?*, *You mean John?*) accompanied by eyebrow *furrows* were more likely to prompt clarification by the speaker, compared to verbal signals accompanied by eyebrow *raises* or no eyebrow movement at all. Crucially, it was also found that eyebrow furrows alone, i.e., without words, were sufficient to occasion clarification by the speaker. Taken together, these results suggest a communicative function of listener eyebrow furrows in signaling "I've *not* received enough information for current purposes" (Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989). Based on these correlational findings we ask: Is there a causal influence of listener eyebrow furrows on speakers' communicative behavior in face-to-face interaction?

To address these questions, we developed a novel experimental paradigm using Virtual Reality technology enabling us to selectively manipulate visual feedback in virtual listeners (see also Chapter 3; Hömke, Holler, & Levinson, submitted). This selective manipulation allowed us to address questions regarding the causal role of eyebrow furrows in interactive face-to-face communication—questions that have previously been impossible to address with such a high degree of experimental control. Participants were asked to have a conversation with different avatars and to answer open questions (e.g., *How was your weekend, what did you do?*). During the participant's answers, the avatar produced different types of visual feedback responses, which were secretly triggered by a confederate. In one condition, the confederate triggered nods in the avatar (baseline 'nod' condition). In a second condition, the confederate triggered nods and, crucially, occasionally an eyebrow furrow instead (experimental 'nod/brow furrow' condition). A control condition was identical to the experimental 'nod/brow furrow' condition except that the occasional eyebrow furrows were replaced with no response at all while the nods were retained (control 'nod/non-response' condition).

If listeners' eyebrow furrowing is irrelevant for the speaker's speaking behavior, one would not expect any differences between the nod condition and the nod/brow furrow condition. However, if listeners' eyebrow furrowing can indeed signal "I've *not* received enough information for current purposes" providing evidence for unsuccessful grounding (Ekman, 1979; Hömke et al., in prep; see Chapter 4), speakers should provide extra information; that is, they should provide longer answers in the nod/brow furrow condition than in the nod baseline condition. Note, however, that rather than providing additional semantic information when talking to an avatar who occasionally furrowed her brows, speakers may produce more hesitations than in the other conditions—unfilled, silent pauses and filled pauses like *uh* and *uhm*—which may alternatively explain any differences in overall answer length. To be able to rule out this possibility, we also measured the frequency and duration of filled pauses and unfilled pauses within each answer.

Speaking behavior, like any other social behavior, varies from individual to individual (Heerey, 2015). In this experiment, two particular individual differences measures of dispositional social sensitivity—the Empathy Quotient (Baron-Cohen & Wheelwright, 2004) and the Fear of Negative Evaluation scale (henceforth 'FNE'; Watson & Friend, 1969)—were hypothesized to modulate the perception of eyebrow movements. Sensitivity to listeners' eyebrow furrows may depend on the speaker's degree of empathy, which is the "drive or ability to attribute mental states to another person/animal, and entails an appropriate affective response in the observer to the other person's mental state" (Baron-Cohen & Wheelwright, 2004, p.168). It has been observed that "to drive your point home in a discussion for far longer than is sensitive to your listener" constitutes low-empathy behavior (Baron-Cohen & Wheelwright, 2004, p.170), suggesting that low-empathy speakers may be less sensitive to listener feedback than high-empathy speakers. To address this issue, participants were asked to complete the Empathy Quotient questionnaire (Baron-Cohen et al., 2004) after the experiment. Sensitivity to listener's eyebrow furrows may also depend on the speaker's degree of 'fear of negative evaluation' (FNE; Watson & Friend, 1969). In contrast with low-FNE individuals, high-FNE individuals are highly concerned with seeking social approval (Watson & Friend, 1969). High-FNE individuals have been shown to exhibit more pro-social behavior (Schlenker, 1980), and to try harder making a good impression during face-to-face conversations (Leary, 1983). According to Leary (1983), "People who are highly concerned about being perceived and evaluated negatively would be more likely to behave in ways that avoid the possibility of unfavorable evaluations and, thus, be more responsive to situational factors relevant to such concerns than individuals who are less apprehensive about others' evaluations of them" (p. 371). One such relevant situational factor may be other's facial expressions. Indeed, high-FNE individuals have been shown to pay more attention to faces (Rossignol, Campanella, Bissot, & Philippot, 2013), particularly to faces expressing negative emotions due to their potentially socially devaluating meaning (Winton, Clark, & Edelmann, 1995; Tanaka & Ikegami, 2015). Since eyebrow furrowing is associated with expressions of negative emotions like anger (Ekman, 1993), one might expect high-FNE individuals to be especially sensitive to listener eyebrow furrows as they occur in the present study. Finally, high-FNE individuals have also been shown to judge their own communicative effectiveness more accurately, that is, in a way that is more consistent with listener's actual understanding, which might be due to their increased sensitivity to listener feedback (Fay, Page, Serfaty, Tai, & Winkler, 2008).

If listeners' eyebrow furrows are not a semiotic, conventional signal but, e.g., a symptom of the listener's cognitive effort (see also Chapter 4), one may expect only high-empathy or high-FNE speakers to be responsive to listeners' eyebrow furrows in the messages they design, due to their stronger social sensitivity. However, if listeners' eyebrow furrows are indeed a semiotic, conventional signal, one may expect all speakers to be sensitive to listeners' eyebrow furrows (although high-empathy or high-FNE speakers might be more so).

The overall aim of the current study was to experimentally test earlier claims based on correlational evidence suggesting that listener eyebrow furrows may serve a communicative function in conversation (see also Chapter 4; Hömke, Holler, & Levinson, in prep). The main hypothesis was that listeners' eyebrow furrows can function as a communicative signal of insufficient understanding, that speakers would produce longer answers in the nod/brow furrow condition than in the nod baseline condition, while individual differences in speakers' social sensitivity may modulate this effect.

Methods

Participants

We recruited 36 native Dutch speakers through the MPI for Psycholinguistics subject database (www.mpi.nl/ppreg) for participation in the experiment. The data of one participant were excluded from all analyses because he provided such long answers to the avatar's questions that we had to interrupt him and end the experiment prematurely in order to be able to test the remainder of the scheduled participants. The data of one additional participant were excluded from all analyses because he excessively looked away from the screen (more often than 2.5 SD above the mean) during avatar listener responses—and therefore he could not have been influenced by differences avatar listener responses. Another participant did not complete the Empathy Quotient questionnaire and was therefore excluded from any analyses including the Empathy Quotient. This resulted in a final sample of 34 participants (18-33 years; mean age = 22.47; 18 females, 16 males), or 33 participants for analyses including the Empathy Quotient (18-33 years; mean age = 22.54; 18 females, 15 males). Each participant was paid €10 and the whole session lasted about one hour.

Apparatus and Materials

Laboratory set-up and equipment. Participants were invited to the Virtual Reality laboratory at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. They were seated in front of a computer screen (HP Compaq LA2405WG) with speakers (Hercules XPS 2.010) wearing a lightweight, head-mounted microphone (DPA-d:fine-88). Audio was recorded using Adobe Audition CS6 and video was recorded using three synchronized video cameras (Sony 3CCD Megapixel) to capture the participant (1) frontally, and (2) laterally, as well as to record a separate computer screen showing exactly what the participant was seeing on their screen (i.e., the avatar). This setup allowed us to link participant and avatar behavior in a time-aligned manner. For each recording session, we synchronized the three videos and the audio file based on audible and visible markers (produced at the beginning of each block) and exported them in Adobe Premier Pro CS6 (MP4, 25 fps). The confederate was seated in the control room next to the experiment room, in front of a keyboard (Apple MB110LL/B) and a computer screen (Acer AL732). The computer screen showed the participant in real time from a frontal view. Audio from the participant's microphone was also transmitted to the control room and played via speakers (Alesis M1Active 520) in real time (see also Procedure).

Avatar characteristics and behaviors. The experiment was programmed in WorldViz's Vizard 5.5 and three different female avatars were created based on a stock avatar produced by WorldViz. Three different female Dutch native speakers were used to pre-record the avatars' speech, which was played at appropriate times during the experiment (one per condition). The avatars' lip movements were programmed to match the amplitude of the pre-recorded speech files (i.e., the higher the amplitude, the wider the avatar opened her mouth), creating an illusion of synchronization. The speech materials consisted of a general introduction (e.g., *Hoi, Ik ben Julia, leuk je te ontmoeten!*; 'Hi, I'm Julia, nice to meet you!' and *Ik heb een aantal vraagen aan jou*; 'I have a couple of questions for you') and a set of 18 open-ended questions (e.g., *Hoe was je weekend, wat heb je allemaal gedaan?*; 'How was your weekend, what did you do?'). The avatar also responded to the participant's answer (e.g., *Oh ja, wat interessant!*; 'Oh, how interesting!') before moving on to the next open question (e.g., *Ik heb nog een vraag aan jou;* 'I have another question for you'), or before closing the interaction (*Hartelijk bedankt voor dit gesprek, ik vond het gezellig!*; 'Thank you very much for this conversation, I enjoyed it!').

The crucial experimental manipulation in the present study was the feedback responses the avatar produced when she was in the listener role (see Fig. 1 for example stills). Critically, these feedback responses were modelled on feedback behavior that occurs in natural conversation and they consisted of head nods (duration of 500 milliseconds from nod onset to nod offset) and in one condition eyebrow furrows (duration of 500 milliseconds from eyebrow furrow onset to eyebrow furrow offset). In the control condition, the avatar produced 'non-responses', periods in which the avatar did not and could not produce any feedback response. That is, during a 'non-response', the avatar was just still (default behavior). Note that the duration of 'non-responses' matched the durations of the other feedback responses precisely (i.e., 500 milliseconds).

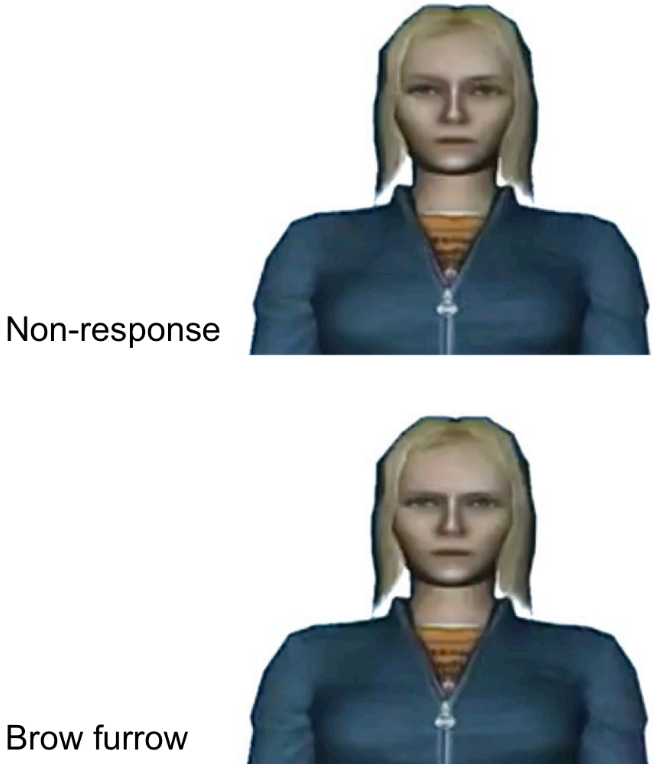

Figure 1. Example stills of a virtual listener producing different types of listener feedback responses (non-response, brow furrow) that were varied across condition.

Questionnaires. The questionnaires consisted of the Dutch version of the Empathy Quotient questionnaire (test-retest reliability: r = 0.97, as reported by Baron-Cohen & Wheelwrigth, 2004) and the Dutch version of the brief 'Fear of Negative Evaluation Scale' (test-retest reliability: r = 0.75, as reported by Leary, 1980). To control for the possibility that any differences in answer length might be driven by differences in perceived naturalness, perceived ease of understanding by the avatar of the participant, and perceived likability of the avatars depending on the different feedback behaviors they produced, we asked participants to fill in three additional questionnaires tapping these three aspects (one for each avatar each participant interacted with, that is, one per listener-feedback condition). The avatar evaluation questionnaires consisted of statements designed to assess the participants' perception of the avatar's (1) humanness (*Ik vond deze avatar menselijk overkomen*; 'This avatar appeared human'), (2) ease of understanding by the avatar of the participant (*Ik denk dat deze avatar mij makkelijk te begrijpen vond*; 'I think this avatar found me easy to understand'), and (3) likability (*Ik vond deze avatar* *sympathiek overkomen*; 'This avatar appeared nice'; *Ik zou vrienden kunnen zijn met deze avatar*; 'I could be friends with this avatar'; *Ik vond deze avatar egocentrisch overkomen*; 'This avatar appeared selfish') as their conversational partner (adapted from Weatherholtz, Campbell-Kibler, & Jaeger, 2014, and the Dutch translations used in the Relationship Questionnaire of Heyselaar, Hagoort, & Segaert 2015). Participants indicated on a 6-point Likert scale their degree of agreement for each statement (1 = *I do not agree at all*, 6 = *I absolutely agree*). Statistical tests confirmed that the perceived humanness, ease of understanding by the avatar of the participant, and likability (rated through scores for niceness, friendship, selfishness, see above) of the avatars did not differ across listener-feedback conditions (see Appendix).

Analysis

Answer length. Answer length was measured in seconds (in ELAN 4.9.3; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), from the first to the last vocalization produced by the speaker in response to each question.

Hesitations. To differentiate changes in answer length due to content from changes in answer length due to hesitations, we measured different types of hesitations, namely the frequency and average duration of filled pauses (*uh*'s and *uhm*'s; Lickley, 2015) and unfilled pauses (audible intra-turn silences longer than 100 milliseconds; see e.g., Eklund, 2004; Lickley, 2015).

Design

We used a within-subject design with avatar listener feedback (nod, eyebrow furrow, non-response) as independent variable and mean answer length as the main dependent variable. Additional dependent variables consisted of the Empathy Quotient (Baron-Cohen & Wheelwrigth, 2004), the Fear of Negative Evaluation score (Leary, 1983), hesitations (frequency and duration of filled and unfilled pauses), as well as the avatar evaluation questionnaire scores assessing perceived humanness, ease of understanding by the avatar of the participant, and likability of each avatar.
The experiment consisted of three blocks, one block per feedback condition (i.e. one per avatar). The set of 18 spoken question stimuli were split up into three sets of 6 questions and each set was assigned to one of the three avatars, meaning each participant heard each question only once. The order of feedback conditions as well

as the assignment of avatars (and thus the 6 questions that were paired with the respective avatars) to the listener feedback conditions was counterbalanced across participants. The order of items within each block was randomized.

Procedure

Participants were seated in front of the computer screen and were asked to meet and have a conversation with three different avatars (see Fig. 1) and to respond to their questions. After a short personal introduction, the avatar asked questions and produced different types of visual feedback responses while participants answered (see Avatar characteristics and behavior). All visual feedback responses of all three avatars were triggered secretly by a confederate, a Dutch native speaker who could see and hear the participant (via a video-camera link), who was blind to the experimental hypotheses (and not informed about the manipulations), and who was instructed to imagine being the actual listener interacting with the participant and to press a button whenever it felt appropriate to provide listener feedback. Which of the confederate's button presses triggered a nod and which a brow furrow (within the nod/brow furrow condition) was varied automatically by the computer program. To avoid unnatural repetitions of eyebrow furrows, we made sure that following each eyebrow furrow, the next one or two feedback responses (randomly varied) would be a nod before a next eyebrow furrow could be produced. Upon each answer completion by the participant, the avatar produced a response to the participant's answer (e.g., 'Oh, how interesting!'), which was also triggered secretly by the confederate. After having finished the conversation with the third avatar, the experiment was over and participants were asked to complete questionnaires before they were debriefed on the purpose of the experiment. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen and informed consent was obtained before and after the experiment.

Statistical Analysis

We used R (R Core Team, 2012) and *lme4* (Bates, Maechler, & Bolker, 2012) to test in a linear mixed-effects model whether answer length differed depending on listener feedback. The initial model was an intercept-only model estimating the mean answer length including intercepts for items (question stimuli) and participants as random effects. Using a likelihood ratio test (using the ‘anova’ function), this intercept model was compared to a model which differed only in that listener feedback (nod, nod/eyebrow furrow, nod/non-response) was included as a fixed effect. To test whether any effect of listener feedback on answer length was modulated by the speakers’ empathy, we first entered listener feedback (nod, nod/eyebrow furrow, nod/non-response) and speaker empathy (EQ score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items (question stimuli) and participants as random effects into the model. This model was then compared to a model that only differed in that listener feedback and speaker empathy were entered as fixed effects *with* interaction term, again using a likelihood ratio test (with the ‘anova’ function). To test whether any effect of listener feedback on answer length was modulated by the speakers’ fear of negative evaluation, we first entered listener feedback (nod, nod/eyebrow furrow, nod/non-response) and fear of negative evaluation (FNE score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items (question stimuli) and participants as random effects into the model. This model was then compared to a model that only differed in that listener feedback and fear of negative evaluation were entered as fixed effects *with* interaction term, again using a likelihood ratio test (with the ‘anova’ function). To test whether any differences in answer length could be explained by differences in hesitations, we subtracted all filled and unfilled pauses—that is, the sum of durations of all filled and unfilled pauses produced within each answer—from the total length of each answer. Then, we ran the same model comparisons again, as described above, with the only difference that the dependent variable now was ‘answer length minus hesitations’.

Results

Speakers’ answer length

Did speakers’ answer length differ depending on listener feedback? As one can see in Figure 2 showing the overall mean answer length by listener feedback condition, speakers indeed produced longer answers in the nod/brow furrow condition than in the nod condition and answers in the control nod/non-response condition were not longer than in the nod condition[20].

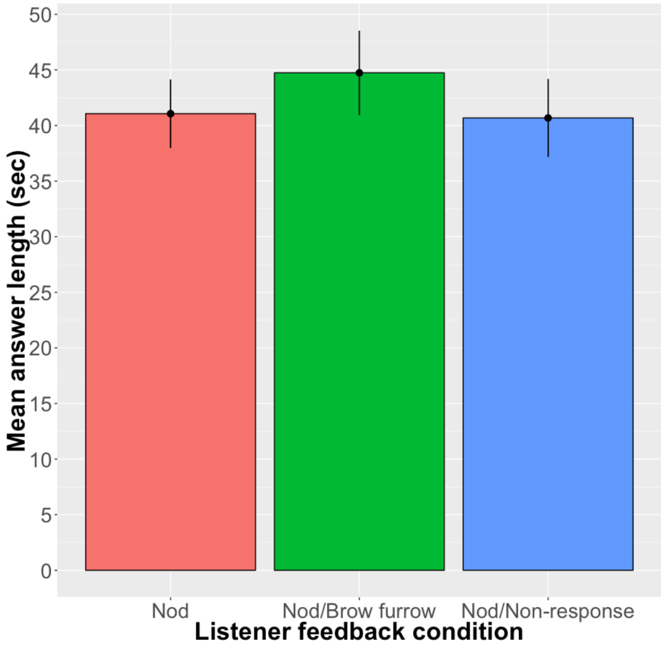

Figure 2. Mean answer length (sec) by listener feedback. Standard errors are represented in the figure by the error bars attached to each column.

To test these predictions statistically, we used R (R Core Team, 2012) and *lme4* (Bates, Maechler, & Bolker, 2012) to examine in a mixed-effects model whether answer length (M = 42.16 sec; SD = 25.05) differed depending on listener feedback. Including 'listener feedback' as fixed effect provided a model with a significantly better fit ($\chi^2(2) = 6.03$, p = .048), revealing that—relative to avatars that nodded throughout (41.4 seconds ± 3.62 [standard error])—the presence of listeners' eyebrow furrows increased speakers' answer length by about 3.77 seconds ± 1.67 standard error (t = 2.25, p = .0246), that is, by approximately eight to eleven words (based on an average of two to three words produced per second in conversation; Levelt, 1999). Also relative to speakers' answer lengths in the nod/non-response control condition

[20] Predicting confederate's feedback button press frequency (number of button presses per answer divided by the length of the same answer in minutes; M = 10.74; SD = 3.52) by feedback condition (nod, nod/brow furrow, nod/non-response), including random intercepts for participants and items, confirmed that button press frequency was consistent across conditions (nod vs. nod/brow furrow: β = -0.05, SE = 0.25, t = -0.255, p = 0.822; nod vs. nod/non-response: β = 0.231, SE = 0.253, t = 0.913, p = 0.362; nod/brow furrow vs. nod/non-response: β = 0.288, SE = 0.253, t = 1.138, p = 0.256).

(41.91 seconds ± 3.61 [standard error]) speakers' answer length was significantly longer in the nod/eyebrow furrow condition ($\beta = 3.34$, $SE = 1.68$, $t = 1.98$, $p = .047$). Speakers' answer lengths in the nod condition and the nod/non-response control condition were statistically indistinguishable ($\beta = 0.43$, $SE = 1.68$, $t = 0.25$, $p = .798$), suggesting that it was not the relatively reduced number of nods in the nod/eyebrow furrow condition that increased the answer length but, as predicted, the presence of eyebrow furrows. Overall, these results support the hypothesis that listener brow furrows can signal "I've *not* received sufficient information for current purposes", such that speakers provide more information, overall resulting in longer answers.

However, there is a possible alternative explanation. Rather than providing additional semantic information, speakers may have produced more hesitations (unfilled, silent pauses and filled pauses like *uh* and *uhm*) when facing an avatar who occasionally furrowed her brows, which may alternatively explain the overall longer answers in the brow furrow condition, compared to the nod condition. To address this issue, we subtracted all filled and unfilled pauses—that is, the sum of durations of all filled and unfilled pauses produced within each answer—from the total length of each answer and then tested again in a linear-mixed effects model whether answer length, now disregarding all filled and unfilled pauses, differed depending on listener feedback. Again, including 'listener feedback' as fixed effect provided a model with a significantly better fit ($\chi^2(2) = 9.38$, $p = .009$), revealing that, relative to avatars that nodded throughout (27.32 seconds ± 2.5 [standard error]), the presence of listeners' eyebrow furrows increased speakers' answer length by about 3.26 seconds ± 1.17 standard error ($t = 2.77$, $p = .005$). Also relative to speakers' answer lengths in the nod/non-response control condition (27.57 seconds ± 2.50 [standard error]) speakers' answer length was significantly longer in the nod/eyebrow furrow condition ($\beta = 3.00$, $SE = 1.18$, $t = 2.54$, $p = .011$). Again, speakers' answer lengths in the nod condition and the nod/non-response control condition were statistically indistinguishable ($\beta = 0.25$, $SE = 1.18$, $t = 0.21$). These results indicate that the observed differences in answer length cannot be explained by differences in hesitations, suggesting that, rather than hesitating more, speakers indeed provided more semantic information when facing an avatar who occasionally furrowed her brows compared to an avatar who nodded throughout, further supporting the hypothesis that listener brow furrows can signal "I've *not* received sufficient information for current purposes", such that speakers provide more information.

Speakers' answer length and individual differences in empathy and fear of negative evaluation

We have seen that speakers provided longer answers when talking to a brow-furrowing listener than when talking to a listener who nodded throughout, and we have also seen that the reason for this was not because they hesitated more when talking to a brow-furrowing listener. Here we investigate whether the effect we found is modulated by individual differences in social sensitivity, focusing on the Empathy Quotient (Baron-Cohen & Wheelwright, 2004) and the Fear of Negative Evaluation scale (Watson & Friend, 1969).

Did the relationship between listener feedback and speakers' answer length depend on speakers' Empathy Quotient? As one can see in Figure 3, high-empathy speakers and low-empathy speakers show similar patterns of results.

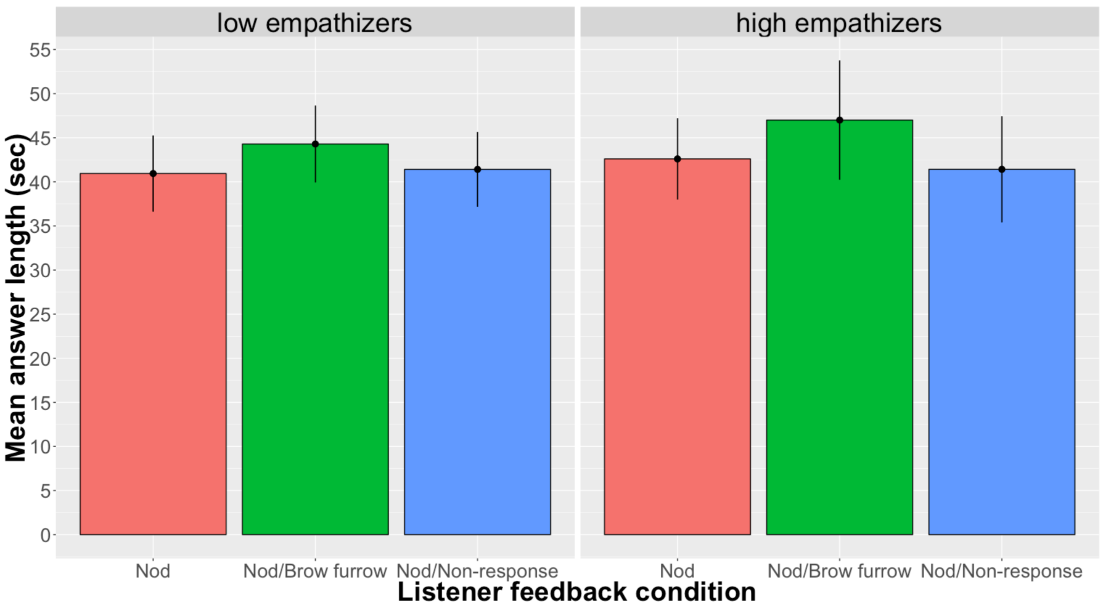

Figure 3. Mean answer length (sec) by listener feedback in low-empathy and high-empathy speakers (median split). Standard errors are represented in the figure by the error bars attached to each column.

We used a mixed-effects model to statistically test whether answer length by listener feedback condition differed depending on the speakers' degree of empathy. We entered listener feedback (nod, nod/brow furrow, nod/non-response) and speaker empathy (EQ score as a scaled and centered continuous variable; M = 43.21; SD =

10.67) as fixed effects (without interaction term), and intercepts for items and participants as random effects into the model. This model was compared to a model that only differed in that listener feedback and speaker empathy was entered as fixed effects *with* interaction term. Including listener feedback and speaker empathy *with* interaction term did not provide a model with a significantly better fit ($\chi^2(2) = 1.77$, $p = .40$), revealing that the effect of listener feedback on speakers' answer length was unaffected by speakers' degree of empathy, also when disregarding filled and unfilled pauses, that is, when predicting 'answer length minus hesitations' ($\chi^2(2) = 1.77$, $p = .32$).

Did the relationship between listener feedback and speakers' answer length depend on speakers' degree of fear of negative evaluation (see Fig. 4)?

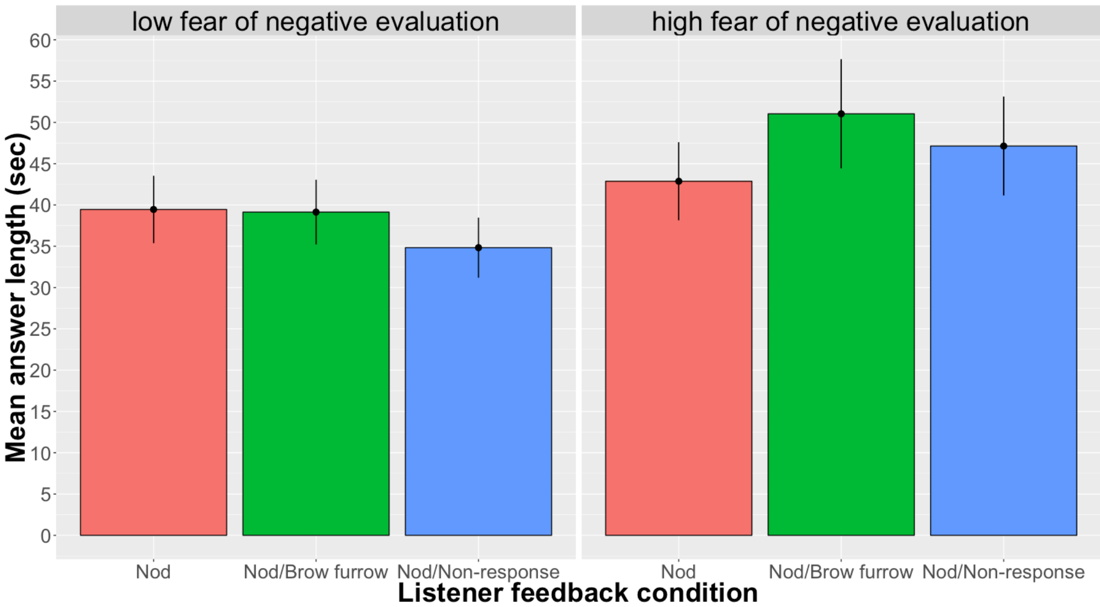

Figure 4. Mean answer length (sec) by listener feedback in speakers with high fear of negative evaluation versus speakers with low fear of negative evaluation (median split). Standard errors are represented in the figure by the error bars attached to each column.

To test this, we entered listener feedback (nod, nod/eyebrow furrow, nod/non-response) and speaker fear of negative evaluation (FNE score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items and participants as random effects into the model. This model was compared to a model that only differed in that listener feedback and speaker fear of negative evaluation was entered as fixed effects *with* interaction term. Including listener feedback and speaker fear of negative evaluation *with* interaction term improved the model fit marginally ($\chi^2(2) = 5.64$, $p = .058$, revealing that the effect of listener feedback on speakers' answer length was not reliably modulated by the speakers' degree of fear of negative evaluation. However, there was a trend in the data revealing that, relative to the nod condition ($\beta = 41.6$, $SE = 3.55$), the higher the speakers' fear of negative evaluation, the longer the answer length in the nod/non-response condition ($\beta = 3.53$, $SE = 1.68$, $t = 2.1$), and the longer the answer length in the nod/brow-furrow condition ($\beta = 3.37$, $SE = 1.67$, $t = 2.01$). Crucially though, the main effect of listener feedback on answer length was still significant ($\beta = 3.66$, $SE = 1.66$, $t = 2.19$) indicating that differences in answer length between the nod and the nod/brow furrow conditions cannot be fully explained by differences in speakers' degree of fear of negative evaluation. Interestingly, however, when predicting 'answer length minus hesitations' (i.e. excluding filled and unfilled pauses from the answer length measure), the marginally significant interaction effect with fear of negative evaluation disappeared. Including listener feedback and fear of negative evaluation *with* interaction term did not provide a model with a significantly better fit ($\chi^2(2) = 3.4$, $p = .18$)[21]. Note that, to explore whether speakers adjusted or marked their speech more locally in response to a listener brow furrow, we also looked at a range of additional variables (speech rate, intensity, pitch change, hesitations), but that none of them explained a significant amount of the data variance (see Appendix for these additional analyses).

Discussion

The central question was: Are speakers sensitive to listener eyebrow furrowing as a communicative signal of insufficient understanding? The findings suggest that they are. In this study, speakers produced longer answers when talking to a brow-furrowing listener than when talking to a listener that nodded throughout, thus supporting the hypothesis that listener eyebrow furrowing can indeed signal

[21] This non-significant interaction of fear of negative evaluation raises the question whether the marginally significant interaction effect of fear of negative evaluation reported above, that is, using answer length including hesitations, actually reflects differences in the amount of semantic information provided, or rather differences in the amount of hesitations produced (see Appendix for additional analyses zooming into this possibility).

insufficient understanding. The observed differences in answer length could neither be alternatively explained by differences in hesitations, nor by differences in speakers' perception of how human or 'natural' the virtual listeners appeared as conversational partners in each listener feedback condition, as assessed by the avatar evaluation questionnaires after the experiment.

However, there is on additional possible alternative explanation to be addressed. One may wonder whether the difference in answer length between the baseline nod condition and the nod/brow furrow condition can really be explained by the absence versus presence of brow furrows. Remember that in the nod/brow furrow condition, a nod was occasionally replaced with a brow furrow, meaning the two conditions did not only differ in the absence versus presence of brow furrows, but also in the overall number of nods. Since nods signal understanding, the relatively reduced overall number of nods in the nod/brow furrow condition rather than the presence of brow furrows may have caused speakers to design longer answers than in the baseline nod condition. If this was the case, one would also have expected longer answers in the control non-response condition than in the baseline nod condition, because the control non-response condition was identical to the nod/brow furrow condition except that the occasional brow furrows were replaced with no response at all (i.e., the control condition differed from the experimental condition only in that brow furrows were absent). However, answer length in the control nod/non-response condition did not differ from answer length in the baseline nod condition. This suggests that the difference in answer length between the nod/brow furrow condition and the nod condition cannot be explained by the reduced number of nods but indeed, as hypothesized, by the presence of eyebrow furrows.

The fact that speakers were influenced by the presence of listener brow furrows does not necessarily entail that these brow furrows are semiotic, conventional signals. In principle, they could be mere symptoms of the listener's cognitive effort. However, we do suggest that listener brow furrowing is not merely a symptom but indeed a conventional signal, treated by the speaker as indicating a need for additional information (see also Chapter 4). One reason for this is that if listener brow furrowing was not a semiotic, conventional signal but a symptom of the listener's cognitive effort, one may have expected only high-empathy or high-FNE speakers to pick up listeners' eyebrow furrows and to be responsive to them in the messages they design, due to their stronger social sensitivity. However, the main effect on answer length was neither modulated by individual differences in speakers' empathy nor by individual differences in speakers' fear of negative evaluation. That is, speakers in general were sensitive to listeners' brow-furrows, leading them to produce longer answers. The lack of empathy/FNE effects may suggest the listener brow furrow is a conventional signal—just as one would not expect high empathy or high FNE to make any difference to the interpretation of a head nod or shake. However, future experimental work is required to provide conclusive insights into the extent to which listener brow furrowing is a communicatively intended, conventional signal.

More generally, the finding has some theoretical implications. We show that listener's facial behavior can shape the speaker's ongoing turn, likely reflecting speaker adjustments at the 'message level' (Levelt, 1989). As such, it provides further support for bilateral accounts of speaking, according to which the listener is an active collaborator coordinating with the speaker moment by moment to maintain mutual understanding. It highlights that speakers in face-to-face communication not only rely on auditory self-monitoring (e.g., Levelt, 1983) but also on visual other-monitoring (see also Clark & Krych, 2004). Although natural human language is multimodal and social-interactive in nature, traditional models of language processing have primarily focused on verbal language and on utterances produced outside of a social-interactive context. This study embraces the multimodal as well as the social-interactive nature of language and it provides further motivation for a paradigm shift, an 'interactive turn' (Kendrick, 2017, p. 7) that is already taking place in psycholinguistics (e.g., Pickering & Garrod, 2004; Levinson, 2016; see also Holler, Kendrick, & Levinson, 2017), but also in the cognitive sciences more generally (Sebanz, Bekkering, & Knoblich, 2006; Jaegher, Paolo, & Gallagher, 2010; Schilbach et al., 2013; Fröhlich et al., 2016).

To conclude, the results suggest that—in addition to visual, emotional, and possible cognitive functions—eyebrow furrowing may serve as a cooperative signal of insufficient understanding. While closing the eyes by blinking may signal "no need to see anymore" because sufficient understanding has been achieved (Hömke, Holler, Levinson, 2017), furrowing the brows—as if trying to see more clearly—appears to signal insufficient understanding (see also Chapter 4; Hömke, Holler, Levinson, in prep), potentially shedding new light on visual origins of mental-state signaling in face-to-face communication.

Acknowledgments: We thank Jeroen Derks and Han Sloetjes for programming support, Ruya Akdag who served as the confederate, and David Peeters, Florian Hintz, Markus Ostarek, the Language & Cognition Department and members of the Virtual Reality Focus Group at the MPI Nijmegen for valuable discussions. This research was funded by the European Research Council (Advanced Grant #269484 INTERACT awarded to Prof. Levinson) and the Max Planck Gesellschaft.

Chapter 5: Appendix

Additional statistical analyses

Speaker's answer length, fear of negative evaluation, and hesitations.

The non-significant interaction of fear of negative evaluation raises the question whether the marginally significant interaction effect of fear of negative evaluation reported above, that is, using answer length including hesitations, actually reflects differences in the amount of verbal information provided, or rather differences in the amount of hesitations produced. Zooming in on this possibility, we tested in a mixed-effects model whether the proportion of filled or unfilled pauses within each answer (that is, the sum of durations of filled or unfilled pauses within each answer divided by the answer's total length) differed depending on listener feedback, and especially whether there was an interaction with fear of negative evaluation. We entered listener feedback (nod, nod/brow furrow, nod/non-response) as a fixed effect, fear of negative evaluation as interaction term, and intercepts for items and participants as random effects into the model. This model was compared to a reduced model without the interaction term of 'fear of negative evaluation' using a Likelihood Ratio Test.

When predicting the proportion of filled pauses per answer, including 'fear of negative evaluation' as interaction term did not improve the model fit significantly ($\chi^2(2) = 2.07$, $p = .35$). Relative to the nod condition ($\beta = 8.06$, $SE = 0.6$), there was neither a significant main effect of listener feedback on the proportion of filled pauses (nod/brow-furrow condition: $\beta = -0.17$, $SE = 0.28$, $t = -0.629$; nod/non-response condition: $\beta = 0.24$, $SE = 0.28$, $t = 0.85$), nor a significant interaction effect of fear of negative evaluation (nod/brow-furrow condition * fear of negative evaluation: $\beta = -0.31$, $SE = 0.28$, $t = -1.098$; nod/non-response condition * fear of negative evaluation: $\beta = -0.38$, $SE = 0.28$, $t = -1.356$).

However, when predicting the proportion of unfilled pauses per answer, including 'fear of negative evaluation' as interaction term did improve the model fit significantly ($\chi^2(2) = 6.68$, $p = .034$). Relative to the nod condition ($\beta = 25.92$, $SE = 1.41$), there was no significant main effect of listener feedback on the proportion of unfilled pauses (nod/brow-furrow condition: $\beta = -0.52$, $SE = 0.74$, $t = -0.701$; nod/non-response condition: $\beta = -0.59$, $SE = 0.74$, $t = -0.796$), but there was a significant interaction effect of fear of negative evaluation, revealing that the higher the speakers' fear of negative evaluation, the higher the proportion of unfilled pauses she produced in the in the nod/non-response condition (β = 1.92, *SE* = 0.75, t = 2.553), relative to the nod condition, but critically, not in the nod/brow furrow condition (β = 0.66, *SE* = 0.75, t = 0.882), relative to the nod condition.

What underlies the FNE-dependent higher overall proportion of unfilled pauses in the nod/non-response condition relative to the nod condition? Does it reflect a longer average duration of unfilled pauses and/or a higher frequency of unfilled pauses? We tested in a mixed-effects model whether the effect of listener feedback on average duration (in milliseconds) or frequency of unfilled pauses (number of unfilled pauses divided by answer length) was modulated by the speakers' degree of fear of negative evaluation. We entered listener feedback (nod, nod/brow furrow, nod/non-response) as a fixed effect, fear of negative evaluation as interaction term, and intercepts for items and participants as random effects into the model. This model was compared to a reduced model without the interaction term of 'fear of negative evaluation' using a Likelihood Ratio Test. When predicting the average duration of unfilled pauses, including 'fear of negative evaluation' as interaction term did not improve the model fit significantly ($\chi^2(2)$ = 0.3, p = .858). However, when predicting the frequency of unfilled pauses, including 'fear of negative evaluation' as interaction term did improve the model fit significantly ($\chi^2(2)$ = 15.89, p = .000), revealing that the higher the speakers' fear of negative evaluation, the higher the frequency of unfilled pauses in the nod/non-response condition (β = 0.26, *SE* = 0.06, t = 3.75) but not in the nod/brow furrow condition (β = 0.04, *SE* = 0.06, t = 0.63), relative to the nod condition (β = 3.79, *SE* = 0.11).

These results suggest that, regarding the difference in answer length between the nod condition and the nod/non-response control condition, the marginally significant interaction effect of fear of negative evaluation reported above (using answer length including hesitations) appears to reflect differences in the amount of hesitations, specifically the frequency of unfilled pauses produced. This suggests that, the higher the speakers' fear of negative evaluation, the more unfilled pauses in the nod/non-response condition relative to the nod condition. However, regarding the difference in answer length between the nod condition and the nod/brow furrow condition—the main contrast of interest—the marginally significant interaction effect of fear of negative evaluation reported above (using answer length including hesitations) indeed appears to reflect differences in the amount of semantic information rather than differences in hesitations. Overall, these results from answer length and hesitations further support the hypothesis that listener brow furrows can signal "I've *not* received sufficient information for current purposes", such that speakers in general provide more verbal information when facing a brow-furrowing avatar listener than when facing an avatar listener that nodded throughout.

Speakers' local adjustments in response to listener brow furrows

We have seen that speakers indeed provided overall longer answers when talking to a listener who occasionally furrowed her brows than when talking to a listener who nodded throughout, suggesting that listener eyebrow furrowing can indeed signal "I've *not* received sufficient information for current purposes". But did speakers adjust or mark their speech more locally in response to a listener brow furrow? Did speakers slow down or speed up, increase or decrease the loudness or the pitch of their speech, did they provide more or less information, or hesitate more or less?

To address this issue, we zoomed in on the nod/brow furrow condition and used several measures comparing speech produced between the onset of a listener nod and the onset of the subsequent listener response (nod segments; n = 865) versus speech produced between the onset of a listener brow furrow and the onset of a subsequent listener response (brow furrow segments; n = 632), resulting in a total of 1497 speech segments. We then tested in linear-mixed effects models whether speech rate (syllables per second), intensity (average, minimum, maximum intensity), pitch change (speaker-specific fundamental frequency minus the median pitch measured over the first 700, 1000, 1500, or 2000 ms of each speech segment), duration without hesitations (subtracting the sum of durations of all filled and unfilled pauses from the total duration of each segment), or the proportion of hesitations (duration of filled or unfilled pauses divided by segment duration) of the speech differed depending on listener feedback (nod segment, brow furrow segment) within the nod/brow furrow condition. We entered listener feedback (nod, brow furrow) as a fixed effect and intercepts for items and participants as random effects into the model. This model was compared to a reduced model without the fixed effect of 'listener feedback' using a Likelihood Ratio Test.

There were no significant main effects of listener feedback. Including 'listener feedback' as fixed effect did not provide a model with a significantly better fit, neither when predicting speech rate ($\chi^2(1)$ = 9e-04, p = .976), nor intensity (average:

$\chi^2(1) = 1.76$, $p = .184$, minimum: $\chi^2(1) = 2.38$, $p = .122$, maximum intensity: $\chi^2(1) = 0.21$, $p = .639$), nor pitch change (first 700ms: $\chi^2(1) = 1.12$, $p = .288$; first 1000 ms: $\chi^2(1) = 0.072$, $p = .787$; first 1500 ms: $\chi^2(1) = 0.1$, $p = .748$; first 2000ms: $\chi^2(1) = 0$, $p = .995$), speech segment duration ($\chi^2(1) = 3.3$, $p = .069$), nor the proportion of hesitations (proportion of filled pauses ($\chi^2(1) = 2.04$, $p = .153$; proportion of unfilled pauses ($\chi^2(1) = 1.79$, $p = .18$).

There were also no significant interaction effects with speakers' fear of negative evaluation. Including 'fear of negative evaluation' as interaction term did not improve the model fit significantly when predicting speech rate ($\chi^2(1) = 0.83$, $p = .364$), intensity (average: $\chi^2(1) = 0.68$, $p = .407$, minimum: $\chi^2(1) = 0.52$, $p = .469$, maximum intensity: $\chi^2(1) = 2.44$, $p = .117$), pitch change (first 700ms: $\chi^2(1) = 2.77$, $p = .09$; first 1000 ms: $\chi^2(1) = 2.57$, $p = .108$; first 1500 ms: $\chi^2(1) = 0.07$, $p = .788$; first 2000ms: $\chi^2(1) = 0.851$, $p = .356$), speech segment duration without hesitations ($\chi^2(1) = 3.3$, $p = .069$), or the proportion of hesitations (proportion of filled pauses ($\chi^2(1) = 3.66$, $p = .055$; proportion of unfilled pauses ($\chi^2(1) = 0.176$, $p = .674$). Thus, within the nod/brow furrow condition, speakers did not change their speech rate, intensity, pitch, the amount of verbal information or hesitations based on whether they received a nod or a brow furrow as listener feedback.

Avatar evaluations: Perceived humanness, ease of understanding by the avatar of the participant, and likability

We have seen that speakers provided longer answers when talking to a brow-furrowing listener than when talking to a listener who nodded throughout, and we have also seen that the reason for this was not because they hesitated more when talking to a brow-furrowing listener. However, perhaps the differences in answer length might be driven by the perceived humanness, ease of understanding by the avatar of the participant, and perceived likability of the avatars as conversational partners in the different listener feedback conditions. To address this issue, we asked participants to fill in three questionnaires tapping these three aspects (see Method).

We tested in linear-mixed effects models whether the scores on each item of the questionnaire (humanness, ease of understanding by the avatar of the participant, and likability [rated through scores for niceness, friendship, selfishness, see Method]) differed depending on listener feedback condition (nod, eyebrow furrow, non-response) and whether this depended on speakers' empathy or fear of negative evaluation. For all models, we entered intercepts for items and participants as random effects. When testing for main effects of ‘listener feedback condition’, we compared a full model including ‘listener feedback condition’ with a reduced model without ‘listener feedback condition’ using a Likelihood Ratio Test. When testing for interaction effects of ‘listener feedback condition’ with speakers’ empathy or fear of negative evaluation, we compared a full model including ‘listener feedback condition’ with empathy or fear of negative evaluation as interaction term with a reduced model without empathy or fear of negative evaluation as interaction term using a Likelihood Ratio Test.

Humanness. Adding listener feedback condition did not improve the model fit of ratings of ‘humanness’ ($\chi^2(2) = 0.35$, $p = .835$; see Figure 6).

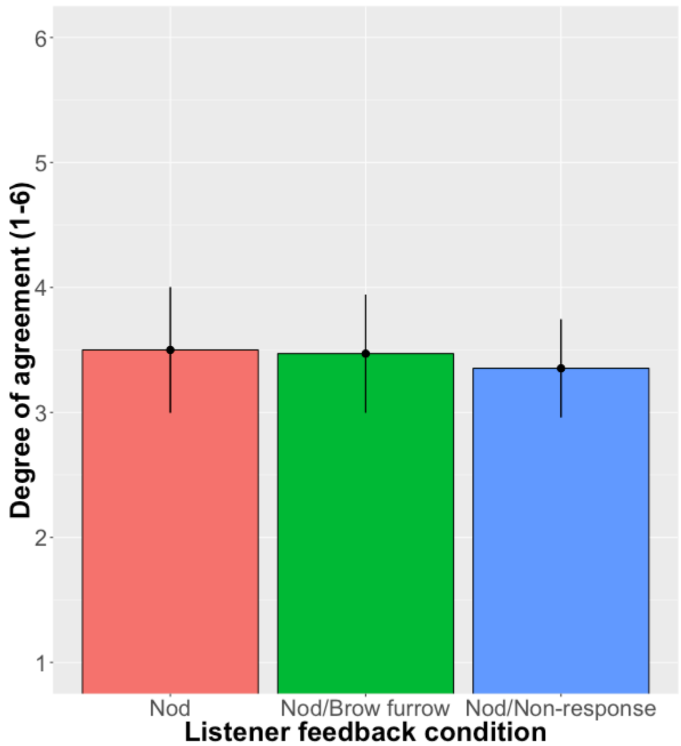

Figure 6. Speakers’ ratings of avatar’s humanness by listener feedback condition.

Also, there were no significant interaction effects, neither for empathy ($\chi^2(2) = 0.059$, $p = .97$) nor for fear of negative evaluation ($\chi^2(2) = 4.76$, $p = .092$). Overall, this suggests that all speakers perceived all three avatars—whether producing nods, brow furrows, or non-responses—as equally human.

Ease of understanding by the avatar of the participant. Adding listener feedback condition did not improve the model fit of ratings of 'ease of understanding by the avatar of the participant' ($\chi^2(2) = 3.64$, $p = .161$). Also, there was no significant interaction effect for empathy ($\chi^2(2) = 3$, $p = .222$). However, there was a significant interaction effect of listener feedback and fear of negative evaluation on ease of understanding by the avatar of the participant ($\chi^2(2) = 8.4$, $p = .014$), revealing that the higher the speakers' fear of negative evaluation, the more they rated the brow-furrowing avatar and the non-response avatar as having difficulty understanding them, relative to the nodding avatar (see Figure 7).

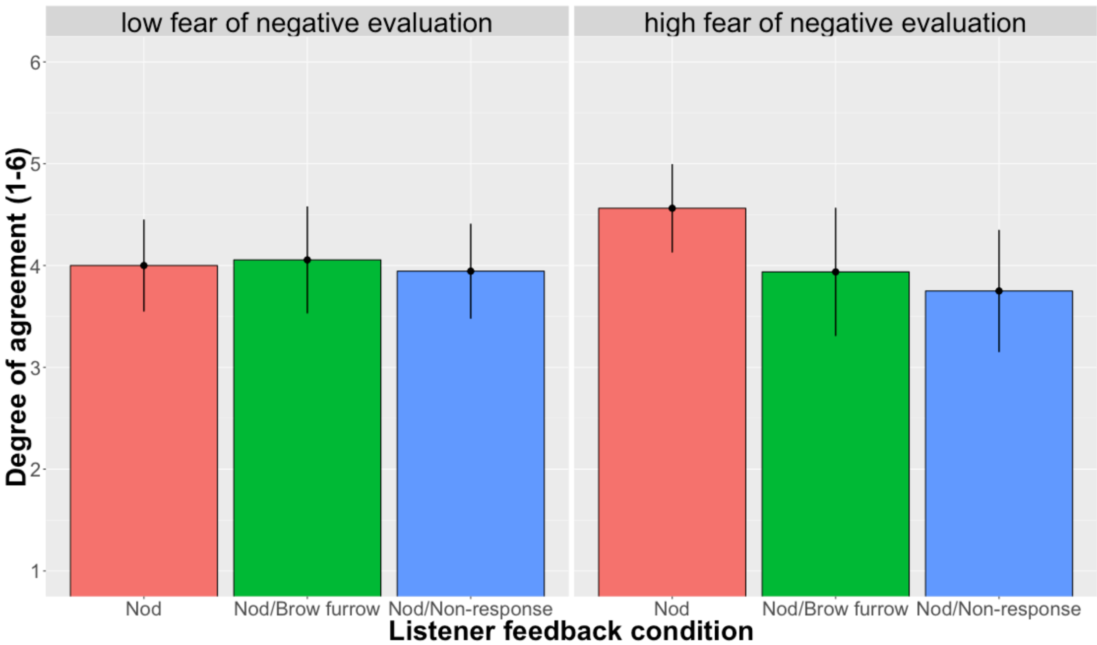

Figure 7. Avatar's perceived ease of understanding by the avatar of the participant as rated by speakers with low- versus high fear of negative evaluation.

Likability. Adding listener feedback condition did not improve the model fit of ratings of 'likability', that is neither of ratings of niceness ($\chi^2(2) = 3.9$, $p = .141$), friendship ($\chi^2(2) = 4.75$, $p = .092$), nor selfishness ($\chi^2(2) = 5.1$, $p = .077$). Also, there were no interaction effects of listener feedback condition and empathy (niceness: $\chi^2(2) = 0.23$, $p = .889$; friendship: $\chi^2(2) = 1.9$, $p = .385$; selfishness: $\chi^2(2) = 0.11$, $p = .994$).

However, there were significant interaction effects of listener feedback condition and fear of negative evaluation (niceness: $\chi^2(2) = 10.26$, $p = .005$; friendship: $\chi^2(2) =$

11.54, p = .003; selfishness: $\chi^2(2)$ = 6.89, p = .031). Note that these significant interaction effects only regard differences between the non-response (control) condition and the nod and the brow furrow condition, respectively. They reveal that the higher the speakers' fear of negative evaluation, the lower the speakers' ratings of the non-response avatar's niceness, friendship potential, and the higher the ratings of the non-response avatar's selfishness, compared to the nodding avatar (niceness: β = -0.78, SE = 0.29, t = -2.7; friendship: β = -0.92, SE = 0.26, t = -3.458; selfishness β = 0.62, SE = 0.23, t = 2.686), as well as compared to the brow-furrowing avatar (niceness: β = -0.87, SE = 0.29, t = -3.012; friendship: β = -0.64, SE = 0.27, t = -2.396; but note the non-significant effect for selfishness: β = 0.27, SE = 0.23, t = 1.173; see Figure 8).

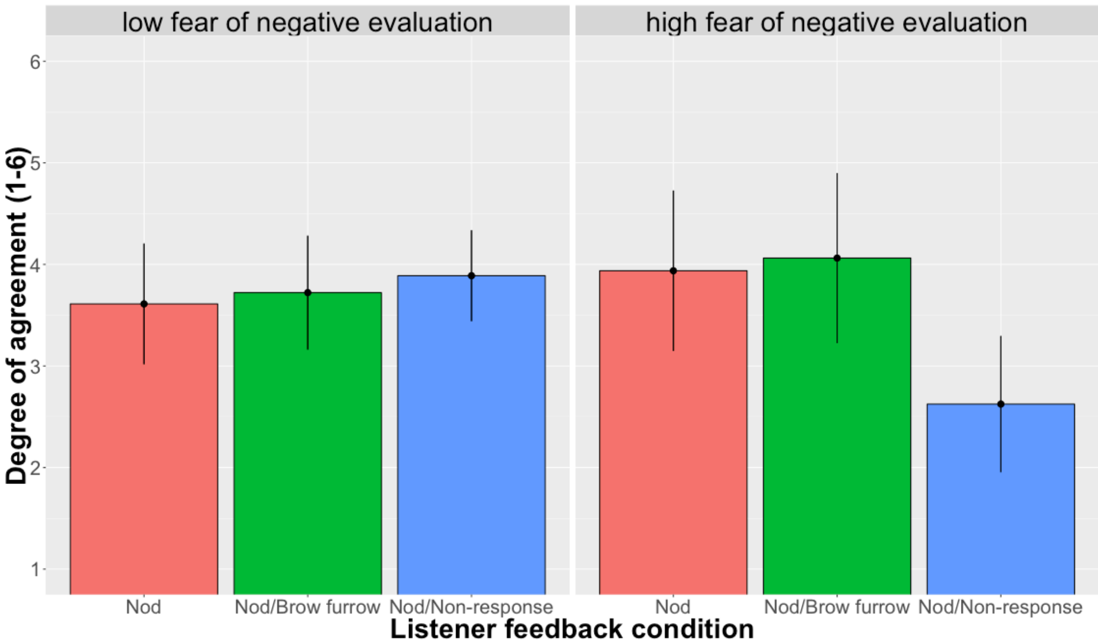

Figure 8 (a). Avatar's perceived niceness as rated by speakers with high- versus low fear of negative evaluation.

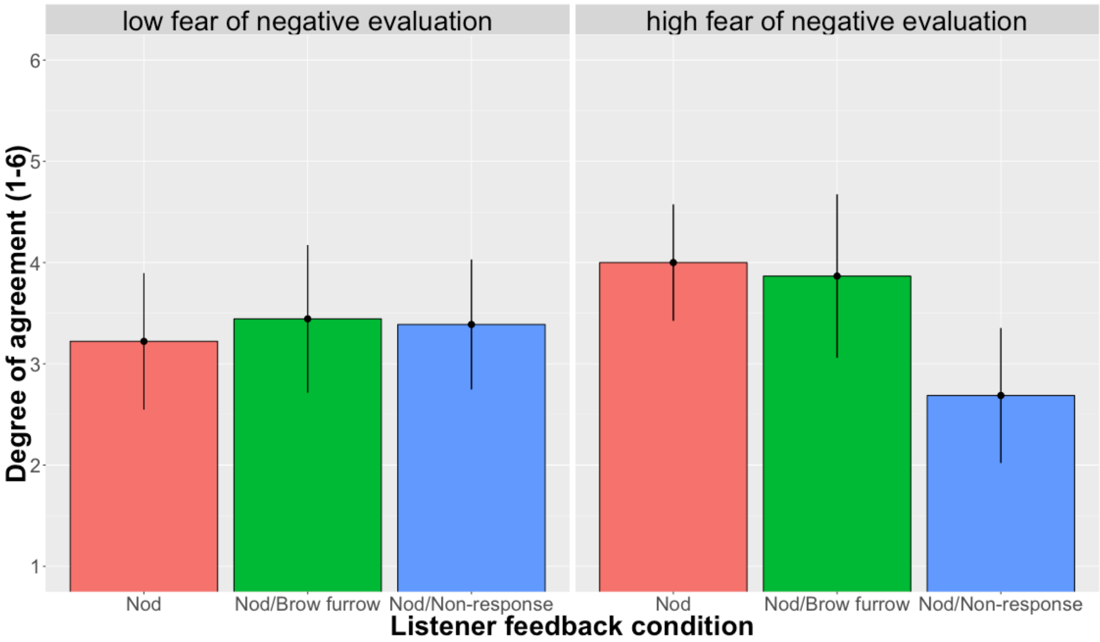

Figure 8 (b). Avatar's perceived friendship potential as rated by speakers with low- versus high fear of negative evaluation.

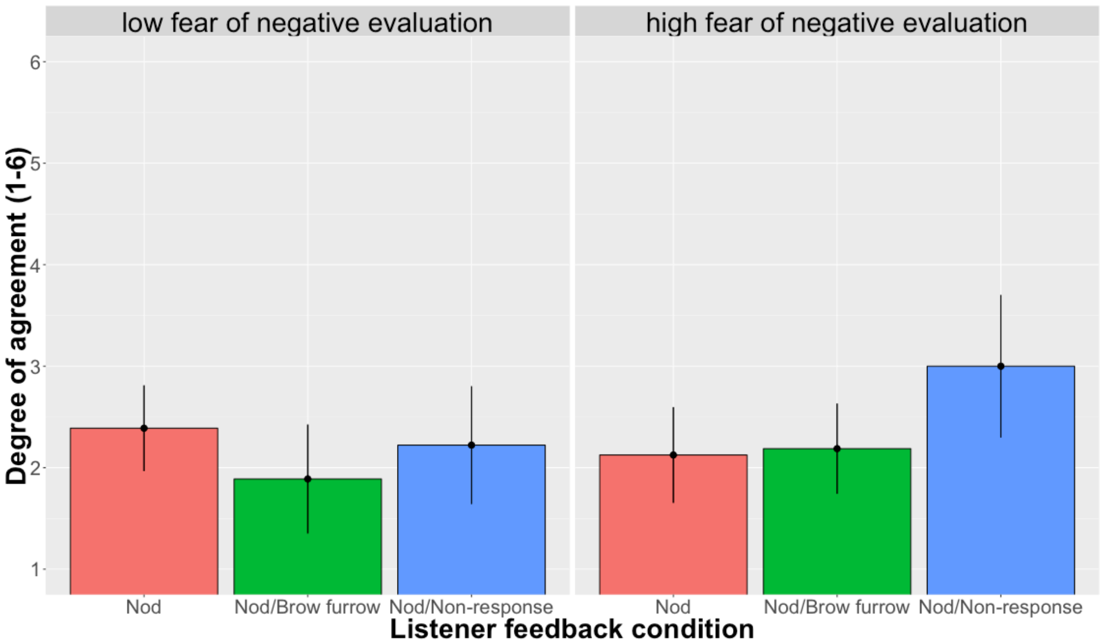

Figure 8 (c). Avatar's perceived selfishness as rated by speakers with low- versus high fear of negative evaluation.

Chapter 6. General discussion

Unlike most other animals, humans tend to face each other in everyday communication. This allows humans to rely not only on vocal but also on various visual bodily behaviors when they communicate (Levinson & Holler, 2014). While the language sciences have made substantial progress in the study of hand gestures (e.g., McNeill, 2000; Kendon, 2004), there is one part of the body that has received relatively little attention despite its omnipresence in and intuitive relevance for everyday face-to-face communication: the face.

The goal of this thesis was to investigate the role of the listener's facial behaviors and to explore their potential communicative signaling functions in managing mutual understanding in face-to-face communication. My general hypothesis was that listeners shape the speaker's ongoing turn through facial feedback signals. On the one hand, some facial behaviors were hypothesized to signal understanding, "I've received enough information for current purposes, please go on", increasing the progressivity of the ongoing turn. On the other hand, other facial behaviors were hypothesized to signal "I've *not* received enough information for current purposes, please clarify", reducing the progressivity of the ongoing turn by inviting additional specification or clarification.

An additional novelty of this thesis lies in the interdisciplinary approach. In an effort to balance ecological validity and experimental control, it combines insights and methodological tools from conversation analysis (qualitative corpus-based), linguistics (quantitative corpus-based), gesture studies (multimodal coding), artificial intelligence (automatic blink detection through facial motion tracking; virtual reality as stimuli), and experimental psychology (controlled experimentation). Making use of this variety of methods, this thesis presents two empirical studies on eye blinking (Chapters 2 and 3) and two empirical studies on eyebrow movements (Chapters 4 and 5). Together, these studies provide converging evidence in support of the hypothesis that listeners' facial feedback signals shape the speaker's speaking[22]. As such, this

[22] Note that the evidence on the tight timing of listener responses in relation to TCU ends (Chapter 2) suggests that not only listener responses shape the speaker's speaking, but also the opposite, that speaker's speaking also shapes listener responses (see also Goodwin, 1980; Healey et al., 2013; Heldner, Hjalmarsson, & Edlund, 2013, on prosodic cues inviting backchannels)—which is what one would expect if "the turn as a unit is interactively

thesis contributes to a better understanding of what it means to speak and listen in face-to-face conversation—the natural habitat of language.

Summary of main findings

Chapter 2 investigated whether eye blinking functions as a type of listener feedback. To explore this possibility, I built a corpus of spontaneous, informal, dyadic Dutch face-to-face conversations, identified short and long listener blinks during extended turns, and measured their occurrence relative to the end of speaking units, that is, relative to the end of turn constructional units (TCUs), the location where feedback typically occurs. Listener blinks were indeed timed to the end of TCUs. Also, long blinks were more likely than short blinks to occur during mutual gaze, with nods or continuers, and their occurrence was restricted to sequential contexts in which signaling understanding was particularly relevant, suggesting a special capacity of long blinks to signal "I've received enough information for current purposes". In the same way in which brow furrowing (as if trying to see more clearly) seems to signal a lack of understanding, closing the eyes by blinking may signal "no need to 'see' anymore" because sufficient understanding has been reached.

Chapter 3 investigated whether speakers are sensitive to (i.e., influenced by) listener blink behavior as a communicative signal. Chapter 3 built on the correlational findings from Chapter 2 and experimentally tested the observationally generated hypothesis that listener blink behavior is taken into account by speakers and that it serves a communicative feedback function signaling "I've received enough information for current purposes". To test this hypothesis, I used virtual reality to develop a novel experimental paradigm enabling us to selectively manipulate blinking in a virtual listener, crucially distinguishing between short and long blinks. I found that speakers unconsciously took into account small differences in listener blink duration, producing shorter answers in the context of long listener blinks, apparently perceiving these as signaling "I've received enough information for current purposes". These findings demonstrate for the first time that, in addition to physiological, perceptual and possible cognitive functions, listener blinking may

---

determined" (Sacks et al., 1974, p. 727), if speaker and listener indeed coordinate moment by moment.

indeed serve as a feedback signal in interactive face-to-face communication—playing a critical role in shaping how we speak, and potentially shedding new light on the visual origins of mental-state signaling.

Chapter 4 addressed the question whether eyebrow raises and furrows play a role in signaling communicative problems. Are eyebrow raises and furrows functionally involved in signaling problems of hearing or understanding? Or might they be mere correlates of verbal signals of problems in hearing or understanding? To address these questions, I collected data from two corpora of face-to-face Dutch conversations, coded the co-occurrence of eyebrow movements with different types of verbal signals of problems in hearing or understanding (or repair initiations), the temporal relationship between the visual and verbal component in these multimodal signals, the type of solutions provided in response, as well as eyebrow movements alone that were treated as signals of problems in hearing or understanding. I found that, while eyebrow raises and furrows co-occurred with all basic types of verbal repair initiations, verbal signals co-occurring with a brow *furrow* were more likely to be responded to with clarifications compared to verbal signals co-occurring with a brow *raise* or no brow movement at all. Second, when speakers were forewarned visually through a brow movement by their recipient that a verbal repair initiations would come up, communicative problems were solved faster than if they were not forewarned through a brow movement, suggesting that brow movements may enhance communicative efficiency. Finally, while brow movements were not necessary for initiating repair, brow furrows alone appeared to be sufficient, suggesting a unique capacity of brow furrows to signal "I've not received enough information for current purposes"—without relying on words. These findings suggest that brow movements are communicative signals in their own right, and that brow raises and furrows may fulfill partially different functions. More generally, they suggest that brow movements go beyond expressing emotions, and that they are frequently used for signaling informational needs in everyday communication.

Chapter 5 investigated whether speakers are sensitive to listener brow furrowing as a communicative signal. It built on the correlational findings from Chapter 4 and experimentally tested the observationally generated hypothesis that listener's eyebrow furrowing is taken into account by speakers and that it serves a communicative feedback function signaling "I've *not* received enough information for current purposes". To test this hypothesis, I used virtual reality to selectively manipulate eyebrow furrowing in a virtual listener (see Chapter 3 for similar methods). I found that speakers produced longer answers when talking to a brow-furrowing listener than when talking to a listener that nodded throughout, thus supporting the hypothesis that listener eyebrow furrowing can indeed signal insufficient understanding. The differences in answer length could neither be alternatively explained by differences in hesitations, nor by differences in speakers' perception of how human or 'natural' the virtual listeners appeared as conversational partners in the different conditions. Taken together, these results suggest that speakers may incorporate listeners' brow behavior into their recipient design, treating listener brow furrows as signaling "I've *not* received sufficient information for current purposes" by providing additional information. Thus, in addition to visual, emotional, and possible cognitive functions, brow furrows may serve as cooperative signals of insufficient understanding.

Theoretical implications and avenues for future research

My findings have a number of theoretical implications. Although natural human language is multimodal and social-interactive in nature, traditional models of language processing have primarily focused on verbal language and on utterances produced outside of a social-interactive context. This thesis embraces the multimodal as well as the social-interactive nature of language and it provides further motivation for a paradigm shift, an 'interactive turn' (Kendrick, 2017, p. 7) that is already taking place in psycholinguistics (e.g., Pickering & Garrod, 2004; Levinson, 2016; see also Holler, Kendrick, & Levinson, 2017), but also in the cognitive sciences more generally (Sebanz, Bekkering, & Knoblich, 2006: Jaegher, Paolo, & Gallagher, 2010; Schilbach et al., 2013; Fröhlich et al., 2016).

The finding that listener blinks and brow movements may be functionally involved in managing mutual understanding highlights the fact that speaking in face-to-face conversation not only involves (auditory) self-monitoring (Levelt, 1983) but also (visual) other-monitoring, which is a missing ingredient of many existing language production models (Clark & Krych, 2004). Although the relevance of recipient design has been acknowledged in language production models (to various extents, see e.g., Keysar et al., 1998; Pickering & Garrod, 2004; Brennan & Hanna, 2009), the empirical focus has been primarily on global information (e.g., what type of addressee am I talking to, what knowledge is mutually shared?) as opposed to local information (e.g., what type of feedback is my addressee currently providing visually?), presumably because global information has been easier to manipulate in controlled experimentation. As Levelt (1989) has already stated, “A speaker, while delivering his utterance, is continuously monitoring himself and his interlocutors, and this feeds back to what he is doing” (p. 8). He further noted that “interlocutors send various signals to the speaker which tell him that something wasn't clear *(eh?),* or that he should go on *(mhm)* (..)” and that “much of this can be done by gaze or gesture” (p.8). The long listener blink and the brow furrow as described in this thesis appear to constitute such gestures, facial gestures signaling successful grounding, in the case of long listener blinks, and signaling unsuccessful grounding, in the case of listener brow furrows. As such, the long listener blink is a type of backchannel (Yngve, 1970), and more specifically, a “facial backchannel” (Bavelas et al., 2014a), while the listener brow furrow could be considered a facial clarification request. Both can provide rapid feedback to the speaker without interrupting the ongoing turn. The findings presented in this thesis provide evidence that a listener’s facial behavior shapes a speaker’s ongoing turn, providing further support for bilateral accounts of speaking, according to which the listener is an active collaborator coordinating with the speaker moment by moment to achieve and maintain mutual understanding.

Future research on language production may investigate in more detail the sensitivity of the language production system to facial listener feedback (especially in term of its form and timing) and illuminate underlying cognitive and neural mechanisms. This may also help distinguishing between different theories of language production. For example, if speakers adapt their production in response to facial addressee feedback with considerable delay, this may support monitoring and adjustment models of language production, potentially pointing to initial “egocentric” production processes (Keysar, Barr, & Horton, 1998). On the other hand, if speakers adapt their production in response to facial addressee feedback flexibly and immediately—“on the fly”, so to say—this may support constraint-based models of language production, pointing to *continuous incrementality* in message preparation, by which “messages can be continually prepared and updated throughout the production process, allowing for fluent production even if new information is added to the message while speaking is underway” (Brown-Schmidt & Konopka, 2015, p. 1).

My findings also have implications for processing models of gesture. Gesture studies have made a substantial progress in understanding hand gestures (e.g., McNeill 2000; Kendon, 2004). However, some influential gesture researchers also include non-manual communicatively intended bodily movements such as facial gestures in their definition of gesture (e.g., Bavelas et al., 2014a). In fact, more and more researchers have turned towards studying facial gesture, for example in the context of depictions, where facial gestures can serve to "stage a scene" (Clark, 2016), for example to impersonate a particular character when telling a story (see also "reenactment", Sidnell, 2006; "facial portrayal", Bavelas, Gerwing, & Healing, 2014b; "multimodal quotation", Stec, Huiskes, & Redeker, 2015), but also in the context of grounding, as studied in this thesis (e.g., Bavelas et al., 2014a). Given the pervasiveness of facial gestures, they may play an important role for language processing in face-to-face conversation. While existing models of gesture production and comprehension have addressed the role of *co-speech* (hand) gesture in language processing (e.g., de Ruiter, 2000; Kita & Özyürek, 2003; Kelly, Özyürek, & Maris, 2010), this thesis may contribute to broadening the scope of gesture research, extending earlier work by Bavelas et al. (2014a) and pointing to a new area of research on *co-listening* (facial) gesture and its role in language processing. Thus, future models of language processing may focus on extending existing models that already embrace the multimodal nature of language (e.g., de Ruiter, 2000; Kita & Özyürek, 2003; Kelly, Özyürek, & Maris, 2010) by taking into account *co-listening* (facial) gesture, and in addition, they may further embrace the social-interactive nature of language. This should also take into account that listeners in interaction not only have to comprehend the incoming signal, predict the speech act, and plan a response while predicting the end (or continuation) of a turn (Levinson, 2016), but that they also have to produce and monitor their own bodily feedback behavior, visually and kinesthetically—all necessary ingredients for attentive listening (or 'doing recipiency') in everyday conversation.

My findings also have implications for conversation analysis and potential methodological fusions of conversation analysis and experimental psychology. While conversation analysis has established a relatively fine-grained classification of different types of listener feedback responses based on the specific functions they fulfill in specific communicative contexts in conversation, it has focused mainly on the verbal modality (Sacks et al., 1974; Schegloff, 1982; Heritage, 1984; Goodwin, 1986; Gardner, 2001). My findings on facial feedback responses extend earlier, conversation analytic work on visual bodily recipient behavior—such as on nods (Stivers, 2008; Whitehead, 2011) and brow movements (Manrique, 2015; Floyd et al., 2016)—and they may further encourage future work in conversation analysis to build and rely on video corpora that allow for the detailed analysis of potentially communicative facial behavior. In addition, my findings may encourage conversation analysts to further consider generating "predictions that could subsequently be tested cumulatively, in ways that contribute to falsifying theories in … psychology" (de Ruiter & Albert, 2017, p. 10). As de Ruiter and Albert (2017) point out, "the fact that preempirical conceptualizations can and often do lead to "shallow" theories that are not firmly grounded in social reality doesn't mean that deriving and testing predictions from theories is in itself a bad idea" (p. 9). Future work in experimental psychology, in turn, may further consider using a conversation analytic approach as a basis for generating experimental hypotheses when studying social interaction. Note, for example, that without my initial reliance on naturalistic data and a conversation analytic approach to studying facial behavior, I would not even have considered listener blinking as a potential type of feedback in spoken Dutch conversation.

Future research on language use across cultures (Sidnell, 2007; Floyd & Dingemanse, 2014) may compare listener 'response systems' across cultures (Levinson & Brown, 2004; see also de Ruiter, 2004), including facial listener feedback. My finding that listener blinking can signal understanding in Dutch is in line with studies on blinking in Yélî Dnye (Levinson, 2015; Levinson & Brown, 2004) and American Sign Language (Sultan, 2004). Similarly, the finding that eyebrow movements can signal unsuccessful grounding in spoken Dutch is in line with examples from other spoken languages like English (Kendrick, 2015), Italian, Chapalaa (Floyd et al., 2014), and Siwu (Dingemanse, 2015), but also with studies on eyebrow movements in signed languages like Dutch Sign Language (Coerts, 1992; De Vos, Van Der Kooi, & Crasborn, 2009) and Argentine Sign Language (Floyd et al., 2014; Manrique, 2016). At least based on this limited number of studies, listener blinking as a signal of successful grounding and listener brow furrowing as a signal of unsuccessful grounding seem to be independent from language modality—since they are used in spoken as well as signed language—as well as from language history—since they have been described in unrelated languages. If listener blinking as a signal of successful grounding and listener brow furrowing as a signal of unsuccessful grounding is shared across a wider range of unrelated languages, it may have evolved due to common pressures of a shared conversational infrastructure (Levinson, 2006; Schegloff, 2006; Dingemanse, Torreira, & Enfield, 2013; Stivers et al., 2009; Levinson, 2016). As such, it would provide further support for the universal pragmatics hypothesis (Levinson, 2000), by which languages vary in the organization of grammar and meaning, while systems of language use are highly similar across cultures.

To conclude, closing the eyelids by blinking (as if having seen enough) and furrowing the eyebrows (as if not seeing clearly) point to a metaphorical use of the muscles surrounding the eyes, signaling sufficient or insufficient understanding. Thus, in everyday social interaction, it is not only the eyes themselves, but crucially, the regions surrounding the eyes, that serve as windows to the mind.

References

Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Atkinson, J. M., Heritage, J. (Eds.) (1984). *Structures of Social action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Augusti, E. M., Melinder, A., and Gredebäck, G. (2010). Look who's talking: pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Frontiers in Psychology.* 1:161.

Baker-Shenk, C. L. (1983). *A microanalysis of the nonmanual components of questions in American Sign Language* (Doctoral dissertation). University of California, Berkeley.

Bangerter, A., & Clark, H. H. (2003) Navigating joint projects with dialogue. *Cognitive Science, 27*, 195-225.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163-175.

Bates D., Maechler M., Bolker B. (2012). lme4: *Linear mixed-effects models using S4 classes*. [WWW document] URL http://CRAN.R-project.org/package=lme4

Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning. An integrated message model of language use in face- to-face dialogue. *Journal of Language and Social Psychology*, *19*, 163–194.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6), 941–952.

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*, *25*(3), 178-198.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13*(1), 63-92.

Bavelas, J., Gerwing, J., & Healing, S. (2014a). Hand and facial gestures in conversational interaction. *The Oxford Handbook of Language and Social Psychology*, 111-130.

Bavelas, J. B., Gerwing, J., & Healing, S. (2014b). Effect of dialogue on demonstrations: Direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Processes, 51,* 619–655.

Beach, W. (1993). Transitional regularities for 'casual' "Okay" usages. *Journal of Pragmatics*, 19, 325–352.

Bolinger, D. L. (1946). Thoughts on 'Yep' and 'Nope.' *American Speech*, 21 (2), 90-95.

Bolinger, D. L. (1983) Intonation and Gesture. *American Speech*, *58*(2), 156-174.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274-291.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, 53, 301–344.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage* (Vol. 4). Cambridge University Press.

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*(3), 833-843.

Bruner, J. (1978). *Learning how to do things with words*. In J. Bruner & A. Garton (Eds.), Human growth and development (pp. 62–84). Oxford, England: Clarendon Press.

Brunner, L. J. (1979). Smiles can be back-channels. *Journal of Personality and Social Psychology*, *37*(5), 728-734.

Casasanto, L. S., Jasmin, K., & Casasanto, D. (2010). Virtually accommodating : Speech rate accommodation to a virtual interlocutor. In R. Camtrabone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. (pp. 127– 132). Portland, OR.

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language*, 92, 234-253.

Coerts, J. (1992). *Nonmanual grammatical markers. An analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands*. Amsterdam: Universiteit van Amsterdam.

Conty, L., Gimmig, D., Belletier, C., George, N., & Huguet, P. (2010). The cost of being watched: Stroop interference increases under concomitant eye contact. *Cognition*, *115*(1), 133-139.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39.

Clark, H. H., & Schaefer, E. R. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*(1991), 127-149.

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567-598.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62-81.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324.

Cummins, F. (2012). Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes*, *27*(10), 1525-1549.

Curl, T. S. (2005). Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions'. *Discourse Processes*, *39*(1), 1-43.

Dachkovsky, S., & Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech*, *52*(2-3), 287–314.

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. In *Psychology of Learning and Motivation* (Vol. 59, pp. 43-95). Academic Press.

Darwin, C. 1872. *The Expression of the Emotions in Man and Animals.* New York: Philosophical Library, 1955.

De Ruiter, Jan Peter (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284–311). Cambridge, UK: Cambridge University Press.

De Ruiter, Jan Peter. 2004. Response systems and signals of recipiency. In Asifa Majid (ed.), Field Manual Volume 9, 53-55. Nijmegen: Max Planck Institute for Psycholinguistics.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515–535.

De Ruiter, J. P., & Albert, S. (2017). An appeal for a methodological fusion of conversation analysis and experimental psychology. *Research on Language and Social Interaction*, *50*(1), 90-107.

De Vos, C., Van Der Kooij, E., & Crasborn, O. (2009). Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and Speech*, *52*(2-3), 315-339.

Dingemanse, M., Torreira, F., & Enfield, N. (2013). Is "huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PloS One*, 8(11), e78273.

Dingemanse M. & Floyd S. (2014). *Conversation across cultures*. In Cambridge Handbook of Linguistic Anthropology (eds Enfield N. J. Kockelman P. & Sidnell J.), pp. 434–464. Cambridge: Cambridge University Press.

Dingemanse M. & Enfield N. J. (2015). Other-initiated repair across languages: towards a typology of conversational structures. *Open Linguistics* 1, 98–118.

Dingemanse M., Kendrick K. H. & Enfield N. J. (2015). A coding scheme for other-initiated repair across languages. *Open Linguistics* 1.

Doane, M. G. (1980). Interaction of eyelids and tears in corneal wetting and the dynamics of the normal human eyeblink. *American Journal of Ophthalmology*, *89*(4), 507-516.

Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science*, *78*(10), 712-725.

Dunbar, R. I. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469-493.

Eibl-Eibesfeidt, I. (1972) Similarities and differences between cultures in expressive movements. In R. A. Hinde (ed.), *Non-verbal communication.* Cambridge: Cambridge Universitv Press.

Ekman, P., & Friesen, W.V. (1975). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues.* Englewood Cliffs, N.J: Prentice-Hall.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P. (1979). About brows: Emotional and conversational signals. *Human Ethology*, 163-202.

Ekman, P. (1985). *Telling lies*. New York, NY: Berkeley Books.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48(4), 384–392.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System*. Salt Lake City, UT: Research Nexus.

Eklund, R. (2004). Disfluency in Swedish human-human and human- machine travel booking dialogues. PhD dissertation, University of Linköping, Sweden.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge, UK: Cambridge University Press.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., et al. (2013). *Huh? What?* – A first survey in 21 languages. In: Hayashi, M., Raymond, G., Sidnell, J., editors. *Conversational Repair and Human Understanding*. Cambridge: Cambridge University Press.

Evinger, C., Manning, K. A., Pellegrini, J. J., Basso, M. A., Powers, A. S., & Sibony, P. A. (1994). Not looking while leaping: the linkage of blinking and saccadic gaze shifts. *Experimental Brain Research*, *100*(2), 337-344.

Fay, N., Page, A. C., Serfaty, C., Tai, V., & Winkler, C. (2008). Speaker overestimation of communication effectiveness and fear of negative evaluation: Being realistic is unrealistic. *Psychonomic Bulletin & Review*, *15*(6), 1160-1165.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Philadelphia, PA: Psychology Press/Taylor & Francis.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication*, *52*(6), 542-554.

Floyd, S., Manrique, E., Rossi, G., & Torreira, F. (2016). The timing of visual bodily behavior in repair sequences: evidence from three languages. *Discourse Process*.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. *Studies in Interactional Sociolinguistics*, *13*, 134-184.

Fox, J., Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., & Pika, S. (2016). Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Scientific Reports*, 6, 25887.

Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. Amsterdam, The Netherlands: John Benjamins.

Garrett, M. F. (1980). Syntactic processes in sentence production. In B. Butterworth (Ed.), *Speech production* (pp. 170– 220). New York: Academic Press.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359-364.

Goodwin, C. (1980). Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inquiry*, *50*(3-4), 272-302.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York, NY: Academic Press.

Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, 9(2–3), 205–217.

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*(1-2), 51-76.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377-388.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics:* 3. *Speech acts.* New York: Academic Press.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279-1296.

Healey, P. G. T., Lavelle, M., Howes, C., Battersby, S., & McCabe, R. (2013). How listeners respond to speaker's troubles. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2506-2511), Berlin: Cognitive Science Society.

Heath, C. (1992). Gesture's discreet tasks: Multiple relevancies in visual conduct and in the contextualization of language. In Peter Auer & Aldo Di Luzio (Eds.), *The Contextualization of Language* (pp. 101–127). Amsterdam: John Benjamins.

Heerey, E. A. (2015). Decoding the dyad: Challenges in the study of individual differences in social behavior. *Current Directions in Psychological Science*, *24*(4), 285-291.

Heldner, M., Hjalmarsson, A., & Edlund, J. (2013). Backchannel relevance spaces. In *Nordic Prosody XI, Tartu, Estonia* (pp. 137-146). Peter Lang Publishing Group.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 299–347). Cambridge, England: Cambridge University Press.

Herrmann, A. (2010). The interaction of eye blinks and other prosodic cues in German Sign Language. *Sign Language and Linguistics*, *13*(1), 3-39.

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods*, *49*(1), 46-60.

Holle, H., & Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior Research Methods*, *47*(3), 837-847.

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522-3536.

Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Frontiers in Psychology*, *6*(98).

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (Eds.). (2016). *Turn-Taking in Human Communicative Interaction*. Lausanne: Frontiers Media.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2017). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*. Advance online publication.

Hömke, P., Holler, J., & Levinson, S. C. (2017). Eye blinking as addressee feedback in face-to-face conversation. *Research on Language and Social Interaction*, *50*(1), 54-70.

Jaegher, H. D., Paolo, E. D., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–447.

Jefferson, G. (1974). Error correction as an interactional resource. Language in Society, 2, 181–199.

Jefferson, G. (1983). *Two explorations of the organization of overlapping talk in conversation: Notes on some orderliness in overlap onset*. Tilburg papers in language and literature, No. 28. Tilburg, NL: Tilburg University, 11–38.

Kaneko, K., & Sakamoto, K. (1999). Evaluation of three types of blinks with the use of electro-oculogram and electromyogram. *Perceptual and Motor Skills*, *88*(3), 1037-1052.

Kaukomaa, T., Peräkylä, A., & Ruusuvuori, J. (2014). Foreshadowing a problem: Turn-opening frowns in conversation. *Journal of Pragmatics*, *71*, 132-147.

Keitel, A., Prinz, W., Friederici, A. D., Von Hofsten, C., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology*, 116(2), 264–277.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260-267.

Kendon, A. (1975). Some functions of the face in a kissing round. *Semiotica*, *15*(4), 299-334.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters* (Vol. 7). Cambridge University Press.

Kendon, Adam. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kendrick, K. H. (2015). The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. *Frontiers in Psychology*, *6*(250), 10-3389.

Kendrick, K. H. (2017). Using conversation analysis in the lab. *Research on Language and Social Interaction*, *50*(1), 1-11.

Kendrick, K. H., & Holler, J. (2017). Gaze direction signals response preference in conversation. *Research on Language and Social Interaction*, 50(1), 12-32.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11(2), 201–258.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*(2), 46-49.

Kita, Sotaro & Asli Özyürek (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48, 16–32.

Kopp, Stefan, Allwood, Jens, Ahlsen, Elisabeth, Grammer, Karl, Stocksmeier, Thorsten, 2008. Modeling embodied feedback in a virtual human. In: Wachsmuth, I., Knoblich, G. (Eds.), *Modeling Communication with Robots and Virtual Humans*, vol. 4930. Springer Verlag, Berlin, pp. 18–37.

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, *47*(7), 567-587.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54-72.

Krahmer, E., & Swerts, M. (2004). More about brows. *From brows to trust: Evaluating embodied conversational agents*, 191-216.

Krauss, R. M., Weinheimer, S., (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology* 4, 343–346.

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, *37*, 1539–1553.

Kraut, R. E., Lewis, S. H., Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology* 43, 718–731.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic books.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371-375.

Lerner, G. H. (1996). On the "semi-permeable" character of grammatical units in conversation: Conditional entry into the turn space of another speaker. *Studies in Interactional Sociolinguistics*, *13*, 238-276.

Lerner, G. H., & Kitzinger, C. (2015). Or-Prefacing in the Organization of Self-Initiated Repair. *Research on Language and Social Interaction*, *48*(1), 58-78.

Lee, D. H., Mirza, R., Flanagan, J. G., & Anderson, A. K. (2014). Optical origins of opposing facial expression actions. *Psychological Science*, *25*(3), 745-752.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223-232.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press, Cambridge.

Levinson, S. C., & Brown, P. (2004, February). *Comparative feedback: Cultural shaping of response systems in interaction.* Talk presented at the workshop on Feedback in Interaction, Nijmegen, The Netherlands.

Levinson, S. C. (2006). On the human "interaction engine". In N. J. Enfield, & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39-69). Oxford: Berg.

Levinson, S. C. (2013). *Cross-cultural universals and communication structures.* In M. A. Arbib (Ed.), Language, music, and the brain: A mysterious relationship (pp. 67-80). Cambridge, MA: MIT Press.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B*, 369 (1651) 20130302.

Levinson, S. C. (2015). Other-initiated repair in Yélî Dnye: Seeing eye-to-eye in the language of Rossel Island. *Open Linguistics*, *1*(1).

Levinson, S. C. (2016). Turn-taking in human communication, origins, and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6-14.

Levinson, S.C. (2017). Natural Forms of Purposeful Interaction among Humans: What Makes Interaction Effective? Unpublished manuscript prepared for the Ernst Strüngmann Forum: *Interactive Task Learning: Agents, Robots, and Humans Acquiring New Tasks through Natural Interactions*, May 21–26, 2017.

Lickley, R. J., (2015). Fluency and disfluency. In: Redford, M.A. (ed) *The Handbook of Speech Production*. Hoboken, NJ: John Wiley, pp. 445-474.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19-44.

Malisz, Z., Włodarczak, M., Buschmeier, H., Skubisz, J., Kopp, S., & Wagner, P. (2016). The ALICO Corpus: Analysing the active listener. *Language Resources and Evaluation*, 50(2), 411–442.

Mandel, A., Helokunnas, S., Pihko, E., & Hari, R. (2014). Neuromagnetic brain responses to other person's eye blinks seen on video. *European Journal of Neuroscience*, *40*(3), 2576-2580.

Mandel, A., Helokunnas, S., Pihko, E., & Hari, R. (2015). Brain responds to another person's eye blinks in a natural setting—the more empathetic the viewer the stronger the responses. *European Journal of Neuroscience*, *42*(8), 2508-2514.

Mandelbaum, J. (2013). Storytelling in conversation. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 492–508). Chichester, England: Wiley-Blackwell.

Manrique, E., & Enfield, N. J. (2015). Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. *Frontiers in Psychology*, *6*, 1326.

Manrique, E. (2016). Other-initiated repair in Argentine Sign Language. *Open Linguistics*, 2, 1-34.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Special issue: Spoken word recognition. *Cognition*, 25(1–2), 71–102.

Mazeland, H. (2007). Parenthetical sequences. *Journal of Pragmatics*, *39*(10), 1816-1869.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32 (7), 855–878.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.

McNeill, D. (Ed.). (2000). *Language and Gesture* (Vol. 2). Cambridge University Press.

Mondada, L. (2011). Understanding as an embodied, situated and sequential achievement in interaction. *Journal of Pragmatics*, *43*(2), 542-552.

Mondada, L. (2014). The temporal orders of multiactivity. In P. Haddington, T. Keisanen, L. Mondada, & M. Nevile (Eds.), *Multiactivity in Social Interaction* (pp. 3–32). Amsterdam, The Netherlands: John Benjamins.

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., & Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proceedings of the Royal Society: Biological Sciences*, 3635-3644.

Nakano, T., & Kitazawa, S. (2010). Eyeblink entrainment at breakpoints of speech. *Experimental Brain Research*, *205*(4), 577-581.

Nakano, T., Kato, M., Morito, Y., Itoi, S., & Kitazawa, S. (2013). Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences*, *110*(2), 702-706.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103.

Nooteboom, S., & Quené, H. (2014). Do speakers try to distract attention from their speech errors?. In Johanneke Caspers, Yiya Chen, Willemijn Heeren, Jos Pacilly, Niels Schiller, and Ellen van Zanten (Eds.). *Above and Beyond the Segments: Experimental Linguistics and Phonetics* (pp. 203-217). Amsterdam: John Benjamins Publishing.

Oster, H. (1978). Facial expression and affect development. In M. Lewis and L. Rosenblum (eds.). *The development of affect.* New York: Plenum.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B*, *369*(1651), 20130300.

Peshori, K. R., Schicatano, E. J., Gopalaswamy, R., Sahay, E., & Evinger, C. (2001). Aging of the trigeminal blink system. *Experimental Brain Research*, *136*(3).

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Ponder, E., & Kennedy, W. P. (1927). On the act of blinking. *Quarterly Journal of Experimental Physiology*, *18*(2), 89-110.

R Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Robinson, J. D. (2014). What "What?" tells us about how conversationalists manage intersubjectivity. *Research on Language and Social Interaction*, *47*(2), 109-129.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147-1153.

Rossano, F., Brown, P., Levinson, S. C. (2009). Gaze, questioning and culture. In: Sidnell, J. (Ed.), *Conversation Analysis: Comparative Perspectives*. Cambridge: Cambridge University Press, pp. 187–249

Rossano, F. (2012). *Gaze behavior in face-to-face interaction* (Unpublished doctoral dissertation). Radboud University Nijmegen, Nijmegen, The Netherlands.

Rossano, F. (2013). Gaze in conversation. In *The Handbook of Conversation Analysis* (Stivers, T. and Sidnell, J., eds), pp. 308–29. Chichester, UK: Wiley-Blackwell.

Rossignol, M., Campanella, S., Bissot, C., & Philippot, P. (2013). Fear of negative evaluation and attentional bias for facial expressions: an event-related study. *Brain and Cognition*, *82*(3), 344-352.

Ryle, G. (1968). 'The Thinking of Thoughts. What is Le Penseur Doing?', reprinted in Ryle (2009) *Collected Papers, Vol. 2*, J. Tanney (ed.) (London: Routledge).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696-735.

Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology*, 15-21. New York, NY: Irvington Publishers.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 361-382.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Washington, DC: Georgetown University Press.

Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. *Studies in Interactional Sociolinguistics*, *13*, 52-133.

Schegloff, E. A. (2004). On dispensability. *Research on Language and Social Interaction*, *37*(2), 95-149.

Schegloff, E. A. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In *Roots of Human Sociality. Culture, Cognition and Human Interaction* (Enfield, N.J. and Levinson, S., eds), pp. 70-96, Oxford: Berg.

Schegloff, E. A. (2007). Conveying who you are: the presentation of self, strictly speaking. In N.J. Enfield & T. Stivers (Eds.), *Person Reference in Interaction: Linguistic, Social, and Cultural Perspectives* (pp. 123–148). Cambridge UniversityPress, Cambridge.

Schegloff, E. A. (2013). Ten operations in self-initiated, same-turn repair. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational Repair and Human Understanding* (pp. 41–70). Cambridge: Cambridge University Press.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4), 393–414.

Schlenker, B. R. (1980). *Impression management*. Brooks/Cole Publishing Company.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70-76.

Selting, M. (2000). The construction of units in conversational talk. *Language in Society*, *29*(04), 477-517.

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends in Cognitive Science.* 13, 127–134.

Shannon, C., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments. *Research on Language and Social Interaction, 39,* 377–409.

Sidnell, J. (2007). Comparative studies in conversation analysis. *Annual Review of Anthropology*, *36*, 229-244.

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, *45*(5), 679-687.

Stec, K., Huiskes, M., & Redeker, G. (2015). Multimodal analysis of quotation in oral narratives. *Open Linguistics, 1,* 531–554.

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language & Social Interaction*, 41(1), 31–57.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10587–10592.

Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and Social Interaction*, *43*(1), 3-31.

Sultan, V. D. (2004). *Blinking as a Recipient Resource in American Sign Language Discourse* (Doctoral dissertation, University of California, Santa Barbara).

Tada, H., Omori, Y., Hirokawa, K., Ohira, H., & Tomonaga, M. (2013). Eye-blink behaviors in 71 species of primates. *PloS ONE*, *8*(5), e66018.

Tanaka, H., & Ikegami, T. (2015). Fear of negative evaluation moderates effects of social exclusion on selective attention to social signs. *Cognition and Emotion*, *29*(7), 1306-1313.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition (2nd ediition): Speech language and communication* (pp. 217–262). San Diego: Academic Press.

Tinbergen, N. (1952). "Derived" activities; their causation, biological significance, origin, and emancipation during evolution. *The Quarterly Review of Biology*, *27*(1), 1-32.

Tomasello, M., Call, J. (2007). *The gestural communication of apes and monkeys*. New York: Lawrence Erlbaum Associates.

Van der Werf, F., Brassinga, P., Reits, D., Aramideh, M., & de Visser, B. O. (2003). Eyelid movements: behavioral studies of blinking in humans under different stimulus conditions. *Journal of Neurophysiology*, *89*(5), 2784-2796.

Van Son, R., Wesseling, W., Sanders, E., & Heuvel, H. (2008). The IFADV corpus: A free dialog video corpus. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, & D. Tapias, (Eds.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)* (pp. 1-8), Marrakech, Morocco.

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., Sebanz, N. (2011). “Making oneself predictable: reduced temporal variability facilitates joint action coordination,” *Experimental Brain Research*, vol. 211, pp. 517-530.

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448.

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change*, *26*(3), 387-420.

Wierzbicka, A. (1991). *Cross-cultural Pragmatics: The Semantics of Human Interaction*. Berlin/New York: Mouton de Gruyter.

Winton, E. C., Clark, D. M., & Edelmann, R. J. (1995). Social anxiety, fear of negative evaluation and the detection of negative emotion in others. *Behaviour Research and Therapy*, *33*(2), 193-196.

Whitehead, K. A. (2011). Some uses of head nods in “third position” in talk-in-interaction. *Gesture*, *11*(2), 103-122.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). “ELAN: A professional framework for multimodality research,” in *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (LREC 2006), Nijmegen, 1556–1559.

Xiong, X., & De la Torre, F. (2013). Supervised descent method and its applications to face alignment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 532-539.

Yngve, V. H. (1970). On getting a word in edgewise. In M. A. Campbell (Ed.), *Papers from the sixth regional meeting, Chicago Linguistics Society* (pp. 567–578). Chicago, IL: Department of Linguistics, University of Chicago.

Zama, A., & Robinson, J. D. (2016). A Relevance Rule Organizing Responsive Behavior During One Type of Institutional Extended Telling. *Research on Language and Social Interaction*, *49*(3), 220-237.

Zametkin, A. J., Stevens, J. R., & Pittman, R. (1979). Ontogeny of spontaneous blinking and of habituation of the blink reflex. *Annals of Neurology*, *5*(5), 453-457.

Zeshan, U. (2004). Interrogative constructions in signed languages: Cross-linguistic perspectives. *Language*, 7–39.

Nederlandse samenvatting

In tegenstelling tot de meeste andere dieren hebben mensen de neiging om elkaar aan te kijken in de dagelijkse communicatie. Hierdoor kunnen mensen niet alleen gebruik maken van vocale maar ook van verschillende visuele lichamelijke gedragingen wanneer ze communiceren. Hoewel de taalwetenschappen aanzienlijke vooruitgang hebben geboekt in onderzoek naar handgebaren, is er een deel van het lichaam dat relatief weinig aandacht heeft gekregen, ondanks zijn alomtegenwoordigheid en intuïtieve relevantie voor dagelijkse communicatie: het gezicht. Het doel van dit proefschrift was om de rol van het gezichtsgedrag van de luisteraar te onderzoeken en om potentiële communicatieve signaalfuncties in kaart te brengen die een rol spelen bij wederzijds begrip in conversatie.

De algemene hypothese was dat luisteraars de aanhoudende beurt van de spreker vormen door gezichtssignalen. Aan de ene kant werden sommige gezichtsgedragingen verondersteld om begrip te signaleren: "Ik heb genoeg informatie gekregen voor de huidige doeleinden, ga alsjeblieft verder", wat de progressiviteit of voorwaartse beweging van de aanhoudende beurt versterkt. Aan de andere kant werden andere gezichtsgedragingen verondersteld te signaleren: "Ik heb *niet* genoeg informatie ontvangen voor de huidige doeleinden, bied alsjeblieft verduidelijking", waardoor de spreker minder geneigd is om verder te gaan doordat er om extra specificatie of verduidelijking wordt gevraagd.

In een poging om balans te brengen tussen ecologische validiteit en experimentele controle, combineert dit proefschrift inzichten en methodologische middelen uit conversatieanalyse (kwalitatief corpusgebaseerd), taalkunde (kwantitatief corpusgebaseerd), gebaarstudies (multimodale codering), kunstmatige intelligentie (automatische detectie van oogknipperen, virtual reality als stimuli) en experimentele psychologie (gecontroleerde experimenten). Gebruikmakend van deze verscheidenheid aan methoden, presenteert dit proefschrift twee empirische studies over oogknipperen (Hoofdstukken 2 en 3) en twee empirische studies over wenkbrauwbewegingen (Hoofdstukken 4 en 5).

Hoofdstuk 2 onderzoekt of oogknipperen kan functioneren als een manier van feedback van luisteraars. Om deze mogelijkheid te onderzoeken, bouwden we een corpus van spontane, informele, dyadische Nederlandse face-to-face gesprekken, identificeerden we kort en lang knipperen van luisteraars tijdens uitgebreide beurten van sprekers en maten hun aanwezigheid ten opzichte van ‘einde van de beurt constructie-eenheden’ (TCU's), momenten waar feedback van de luisteraar typisch optreedt. Het knipperen van luisteraars was inderdaad getimed tot het einde van TCU's. Ook was lang knipperen waarschijnlijker dan kort knipperingen tijdens wederzijdse blik, met knikken of ‘continuers’, en hun optreden was beperkt tot communicatieve contexten waarin het signaleren van begrip bijzonder relevant was. Dit suggereert dat lang knipperen een speciale capaciteit heeft om te signaleren: "Ik heb voldoende informatie ontvangen voor de huidige doeleinden". Op dezelfde manier als waarop het fronsen van de wenkbrauwen (alsof men duidelijker probeert te zien) een gebrek aan begrip lijkt te signaleren, kan het sluiten van de ogen door knipperen een signaal zijn van “niet meer hoeven ‘zien’ ” omdat voldoende begrip bereikt is.

Hoofdstuk 3 onderzoekt of sprekers gevoelig zijn voor knippergedrag van luisteraars als een communicatief signaal; of het taalgedrag van de spreker wordt beïnvloed door het knippergedrag van de luisteraar in face-to-face communicatie. Hoofdstuk 3 bouwt voort op de correlationele bevindingen uit hoofdstuk 2 en test op een experimentele manier de observationeel gegenereerde hypothese dat sprekers rekening houden met het knippergedrag van de luisteraar en dat het een communicatieve feedbackfunctie heeft en signaleert dat de luisteraar voldoende informatie heeft ontvangen. Om deze hypothese te testen, hebben we Virtual Reality gebruikt om een nieuw experimenteel paradigma te ontwikkelen waarmee we selectief het knipperen in een virtuele luisteraar kunnen manipuleren, waarbij we een cruciaal onderscheid maakten tussen kort en lang knipperen. We ontdekten dat sprekers onbewust rekening hielden met korte verschillen in de knipperduur van luisteraars, kortere antwoorden produceerden in de context van lang knipperen van luisteraars die kennelijk werden waargenomen als "Ik heb genoeg informatie ontvangen voor huidige doeleinden". Onze bevindingen tonen voor het eerst aan dat het knipperen van luisteraars, naast fysiologische, perceptuele en mogelijke cognitieve functies, inderdaad kan dienen als een feedbacksignaal in interactieve face-to-face communicatie.

Hoofdstukken 2 en 3 onderzoeken het knipperen van het oog van de luisteraar als een mogelijk signaal van *begrip*. De volgende hoofdstukken, hoofdstuk 4 en 5, richten zich ook op het gebied van de ogen van de luisteraar en onderzoeken wenkbrauwbewegingen als mogelijke signalen van *onbegrip*.

Hoofdstuk 4 onderzoekt de rol van wenkbrauwbewegingen bij het signaleren van communicatieve problemen. Zijn het optrekken en fronsen van de wenkbrauwen functioneel betrokken bij het signaleren van problemen in gehoor of begrip of zouden ze epifenomenaal kunnen zijn, louter correlaten van verbale signalen van problemen bij horen of begrijpen? Om deze vragen te beantwoorden, verzamelden we gegevens van twee corpora van face-to-face Nederlandse gesprekken, codeerden we het samen voorkomen van wenkbrauwbewegingen met verschillende soorten verbale signalen van problemen bij horen of begrijpen (of reparatie-initiaties), de temporele relatie tussen de visuele en verbale component in deze multimodale signalen, het soort oplossingen dat in reactie daarop wordt gegeven en alleen wenkbrauwbewegingen die worden behandeld als signalen van problemen met horen of begrijpen. We ontdekten dat, terwijl het optrekken van wenkbrauwen en fronsen samen met alle basistypen van verbale reparatieinitiatieven voorkwamen, de kans hoger was dat verbale signalen die samen met fronsen werden geproduceerd door een verhelderingen werden opgevolgd dan verbale signalen die samen met opgetrokken wenkbrauwen werden geproduceerd of verbale signalen met helemaal geen wenkbrauwbeweging. Ten tweede, toen sprekers door de luisteraar vooraf visueel gewaarschuwd werden dat er een verbale reparatie-initiatie zou komen, waren communicatieve problemen sneller opgelost dan wanneer ze niet vooraf werden gewaarschuwd door een wenkbrauwbeweging, wat suggereert dat wenkbrauwbewegingen de communicatieve efficiëntie kunnen verbeteren. Hoewel wenkbrauwbewegingen niet noodzakelijk waren om reparatie te initiëren, leken wenkbrauwen wel voldoende te zijn, wat suggereert dat fronsen een speciale capaciteit heeft om te signaleren: "Ik heb *niet* genoeg informatie ontvangen voor huidige doeleinden" - zonder afhankelijk te zijn van woorden. Deze bevindingen suggereren dat wenkbrauwbewegingen op zichzelf communicatieve signalen zijn en dat het optrekken van wenkbrauwen en fronsen mogelijk gedeeltelijk verschillende functies vervullen. Meer in het algemeen suggereren ze dat het optrekken van wenkbrauwen en fronsen verder gaan dan het uiten van emoties (bijvoorbeeld verrassing of woede) en dat ze vaak worden gebruikt voor het signaleren van informatiebehoeften in de dagelijkse communicatie.

Hoofdstuk 5 onderzoekt of sprekers gevoelig zijn voor het fronsen van de luisteraar als een communicatief signaal. Hoofdstuk 5 bouwt voort op de correlationele bevindingen uit hoofdstuk 4 en onderzoekt op een experimentele manier de observationeel gegenereerde hypothese dat het fronsen van de luisteraar door de sprekers in aanmerking wordt genomen en dat het een communicatieve feedbackfunctie dient die aangeeft "Ik heb *on*voldoende informatie ontvangen voor de huidige doeleinden". Om deze hypothese te testen, gebruikten we Virtual Reality om selectief fronsen in een virtuele luisteraar te manipuleren (zie hoofdstuk 3 voor vergelijkbare methoden). We ontdekten dat sprekers langere antwoorden produceerden wanneer ze met een fronsende luisteraar praatten dan wanneer ze met een luisteraar praatten die alleen maar knikte. Dit ondersteunt onze hypothese dat het fronsen van de luisteraar inderdaad onvoldoende begrip kan signaleren. De verschillen in antwoordlengte konden evenmin worden verklaard door verschillen in aarzeling, noch door verschillen in de perceptie van sprekers over hoe menselijk of 'natuurlijk' de virtuele luisteraars als gesprekspartners overkwamen in de verschillende condities. Al met al suggereren onze resultaten dat sprekers het wenkbrauwgedrag van luisteraars opnemen in het ontwerp van hun ontvanger, terwijl ze het fronsen van de luisteraar interpreteerden als "Ik heb onvoldoende informatie voor huidige doeleinden ontvangen" door aanvullende informatie te verstrekken. Onze bevindingen tonen aan dat fronsen, naast visuele, emotionele en mogelijke cognitieve functies, kan dienen als coöperatief signaal van onvoldoende begrip. Net als onze bevindingen over knipperen van de ogen, lichten onze bevindingen over fronsen het belang van de oogregio voor het signaleren van de mentale toestand toe, een cruciaal ingrediënt voor het bereiken van intersubjectiviteit in de dagelijkse communicatie.

Tezamen ondersteunen de onderzoeken beschreven in dit proefschrift onze hypothese dat signalen uit het gezicht van luisteraars het spreken van de spreker beïnvloeden. Als zodanig draagt het bij aan een beter begrip van spreken en luisteren in face-to-face gesprekken - de natuurlijke habitat van taal.

English summary

Unlike most other animals, humans tend to face each other in everyday communication. This allows humans to rely not only on vocal but also on various visual bodily behaviors when they communicate. While the language sciences have made substantial progress in the study of hand gestures, there is one part of the body that has received relatively little attention despite its omnipresence in and intuitive relevance for everyday face-to-face communication: the face.

The goal of this thesis was to investigate the role of the listener's facial behaviors and to explore their potential communicative signaling functions in managing mutual understanding in face-to-face communication. The general hypothesis was that listeners shape the speaker's ongoing turn through facial feedback signals. On the one hand, some facial behaviors were hypothesized to signal understanding, "I've received enough information for current purposes, please go on", increasing the forward movement of the ongoing turn. On the other hand, other facial behaviors were hypothesized to signal "I've *not* received enough information for current purposes, please clarify", reducing the forward movement of the ongoing turn by inviting additional specification or clarification.

In an effort to balance ecological validity and experimental control, this thesis combines insights and methodological tools from conversation analysis (qualitative corpus-based), linguistics (quantitative corpus-based), gesture studies (multimodal coding), artificial intelligence (automatic blink detection through facial motion tracking; virtual reality as stimuli), and experimental psychology (controlled experimentation). Making use of this variety of methods, this thesis presents two empirical studies on eye blinking (Chapters 2 and 3) and two empirical studies on eyebrow movements (Chapters 4 and 5).

Chapter 2 investigates whether eye blinking might function as a type of listener feedback. To explore this possibility, I built a corpus of spontaneous, informal, dyadic Dutch face-to-face conversations, identified short and long listener blinks during extended turns, and measured their occurrence relative to the end of speaking units, that is, relative to the end of turn constructional units (TCUs), the location where feedback typically occurs. Listener blinks were indeed timed to the end of TCUs. Also, long blinks were more likely than short blinks to occur during mutual gaze, with nods or continuers, and their occurrence was restricted to communicative

contexts in which signaling understanding was particularly relevant, suggesting a special capacity of long blinks to signal “I’ve received enough information for current purposes”. In the same way in which brow furrowing (as if trying to see more clearly) seems to signal a lack of understanding, closing the eyes by blinking may signal “no need to ‘see’ anymore” because sufficient understanding has been reached.

Chapter 3 investigates whether speakers are sensitive to listener blink behavior as a communicative signal, that is, whether the speaker’s linguistic behavior is influenced by listener’s blink behavior in face-to-face communication. Chapter 3 builds on the correlational findings from Chapter 2 and experimentally tests the observationally generated hypothesis that listener blink behavior is taken into account by speakers and that it serves a communicative feedback function signaling “I’ve received enough information for current purposes”. To test this hypothesis, I used virtual reality to develop a novel experimental paradigm enabling us to selectively manipulate blinking in a virtual listener, crucially distinguishing between short and long blinks. I found that speakers unconsciously took into account small differences in listener blink duration, producing shorter answers in the context of long listener blinks, apparently perceiving these as signaling “I’ve received enough information for current purposes”. These findings demonstrate for the first time that, in addition to physiological, perceptual and possible cognitive functions, listener blinking may indeed serve as a feedback signal in interactive face-to-face communication—playing a critical role in shaping how we speak, and potentially shedding new light on the visual origins of mental-state signaling.

Chapters 2 and 3 investigate listener’s eye blinking as a potential signal of understanding. The next chapters, Chapters 4 and 5, also focus on the eye region of the listener’s face, investigating eyebrow movements as potential signals of non-understanding.

Chapter 4 investigates the role of eyebrow movements in signaling communicative problems. Are eyebrow raises and furrows functionally involved in signaling problems of hearing or understanding or might they be epiphenomenal, mere correlates of verbal signals of problems in hearing or understanding? To address these questions, I collected data from two corpora of face-to-face Dutch conversations, coded the co-occurrence of eyebrow movements with different types of verbal signals of problems in hearing or understanding (or repair initiations), the temporal relationship between the visual and verbal component in these multimodal signals, the type of solutions provided in response, and eyebrow movements alone that were treated as signals of problems in hearing or understanding. I found that, while eyebrow raises and furrows co-occurred with all basic types of verbal repair initiations, verbal signals co-occurring with a brow furrow were more likely to be responded to with clarifications compared to verbal signals co-occurring with a brow raise or no brow movement at all. Second, when speakers were forewarned visually through a brow movement by their recipient that a verbal repair initiations would come up, communicative problems were solved faster than if they were not forewarned through a brow movement, suggesting that brow movements may enhance communicative efficiency. Finally, while brow movements were not necessary for initiating repair, brow furrows alone appeared to be sufficient, suggesting a unique capacity of brow furrows to signal "I've not received enough information for current purposes"—without relying on words. These findings suggest that brow movements are communicative signals in their own right, and that brow raises and furrows may fulfill partially different functions. More generally, they suggest that brow raises and furrows go beyond expressing emotions (e.g., surprise or anger) and that they are frequently used for signaling informational needs in everyday communication.

Chapter 5 investigates whether speakers are sensitive to listener brow furrows as a communicative signal. It builds on the correlational findings from Chapter 4 and experimentally tests the observationally generated hypothesis that listener's eyebrow furrowing is taken into account by speakers and that it serves a communicative feedback function signaling "I've *not* received enough information for current purposes". To test this hypothesis, I used virtual reality to selectively manipulate eyebrow furrowing in a virtual listener (see Chapter 3 for similar methods). I found that speakers produced longer answers when talking to a brow-furrowing listener than when talking to a listener that nodded throughout, thus supporting our hypothesis that listener eyebrow furrowing can indeed signal insufficient understanding. The differences in answer length could neither be alternatively explained by differences in hesitations, nor by differences in speakers' perception of how human or 'natural' the virtual listeners appeared as conversational partners in the different conditions. Taken together, our results suggest that speakers incorporate listeners' brow behavior into their recipient design, treating listener brow furrows as signaling "I've *not* received sufficient information for current purposes" by providing additional information. Our findings demonstrate that, in addition to visual, emotional, and possible cognitive functions, brow furrows may serve as cooperative signals of insufficient understanding. Like our findings on eye blinking, our findings on eyebrow furrowing highlight the importance of the eye region for mental-state signaling, a crucial ingredient for achieving intersubjectivity in everyday communication.

Together this thesis provides converging evidence in support of our hypothesis that listeners' facial feedback signals shape the speaker's speaking. As such, it contributes to a better understanding of what it means to speak and listen in face-to-face conversation—the natural habitat of language.

Acknowledgements

I was a second-year psychology student sitting in a boring meeting of our housing association when I met Lilla Magyari. She had just started her PhD and it was her who first told me about Steve Levinson and the fascinating and adventurous cross-linguistic research being done at the Language and Cognition Department of the Max Planck Institute in Nijmegen. All of a sudden, the meeting of the housing association had become much less boring. This thesis marks the end of my PhD, but it also marks the end of about 11 years working at the Language and Cognition Department of the Max Planck Institute in Nijmegen: First as a student assistant, then as a bachelor's student, a master's student, a research assistant, and eventually as a PhD student. It has been a great privilege for me to be part of this diverse and intellectually stimulating environment for such a long time.

I am deeply grateful to my strikingly complementary PhD supervisors—Judith Holler and Steve Levinson—who have supported me throughout my PhD project with their knowledge and experience, while allowing me plenty of freedom to pursue my own ideas and interests. To Judith, thank you for being a very attentive and knowledgeable mentor and a brilliant teacher with admirable problem-solving skills and important attention for detail. Steve, thank you for sharing your wisdom and vision with me, for thinking outside the box, for being a constant source of inspiration, and for spreading genuine passion. It was a huge privilege for me to work with the two of you.

I would like to thank Lilla Magyari for first telling me about Steve Levinson and his research and also for getting me in touch with Asifa Majid, who hired me as a student assistant shortly after that. I would also like to thank all my other colleagues at the Language and Cognition Department, from whom I have received much useful input, and in particular Nick Enfield, Gunter Senft, Kobin Kendrick, Sean Roberts, Tomas Lehecka, Sara Bögels, Marisa Casillas, and Mark Dingemanse. I extend special thanks to Kobin Kendrick and also Elliott Hoey for introducing me to the world of Conversation Analysis. I very much enjoyed our data sessions and I learned a lot about the value of zooming in on the details of what people actually do in everyday conversation.

I am grateful to my fellow students and friends at the Language and Cognition Department who have shared the PhD adventure with me: Julija Baranova, Mathias

Barthel, Luis Miguel Berscia, Kang-Suk Byun, Jeremy Collins, Rebecca Defina, Gabriela Garrido, Rosa Gisladottir, Elliott Hoey, Elisabeth Manrique, Tayo Neumann, Giovanni Rossi, Sebastian Sauppe, Emma Valtersson, and Ewelina Wnuk. I extend special thanks to my lovely, and wonderfully diverse office mates: Elliott Hoey, Ewelina Wnuk, and Emma Valtersson. Elliott, "A conversation analyst and a psychologist walk into a bar…" could be the beginning of a joke or the beginning of a story about Elliott and Paul. I enjoyed chatting with you about social behavior in public places (e.g., bars) and the often illuminating differences in perspectives we had. Ewelina, it was fun to make a list of different ways to greet in different languages on our whiteboard and I enjoyed our sometimes more and sometimes less profound conversations in English, Spanish, Dutch, or a mix of these languages. Emma, thanks for helping me with your programming skills many times and sometimes with delicious homemade Swedish cinnamon buns. I am glad you still live in Nijmegen. Thank you all for your companionship. I really hope we will continue seeing each other in the future.

I would also like to thank David Peeters and Evelien Heyselaar for introducing me to Virtual Reality, Renske Schilte for helping me build the video corpus and for serving as a confederate in one of my experiments, as well as my excellent interns, Ruya Akdag (who also served as a confederate), Kim Koopmans, and Marina Koleva, for helping me code. I express my gratitude to the Technical Group, in particular to Nick Wood (who will sadly not be able to read this) and Jeroen Geerts for teaching me how to improve the way to record and process audio and video data, Jeroen Derks for programming support for my VR experiments, Ronald Fischer for helping me with the eye-tracking glasses, and Peter Nijland for great Mac and Photoshop support (see cover). Thanks to The Language Archive, especially to Binyam Gebrekidan Gebre for the programming required for automatic blink detection and Han Sloetjes for very helpful ELAN support. I am grateful to Edith Sjoerdsma, the IMPRS office, Els den Os, Dirkje van de Aa, and Kevin Lam, the administration, and the library. Our incredible "BibliotheKarin" Kastens managed to get me a printed copy of a master's thesis on blinking in American Sign Language (library loan from a US university)—in the blink of an eye.

Much appreciation is also owed to the feedback and insight I have received from discussions with Herb Clark, Lorenza Mondada, and Wendy Sandler, as well as the members of the Dialogue Project at the MPI Nijmegen. I would also like to thank several scholars who have inspired and supervised me prior to my PhD. Ton Dijkstra, for the enthusiasm you spread about research on language and communication in an undergraduate psychology course and for paving the way for my internship with Jonathan Grainger at the CNRS in Marseille. Asifa Majid, for first introducing me to the Language and Cognition Department by hiring me as a student assistant, and later, for always challenging me when supervising my bachelor's thesis and co-supervising my master's thesis. Lera Boroditsky, for accepting me as an intern in your lab at Stanford, for co-supervising me as a master's student, for teaching me how to think beyond (e.g., disciplinary) boundaries, and more generally, for being a great source of inspiration as scientist and human being. In addition to Steve Levinson, I would like to thank Tanya Stivers, Nick Enfield, Mark Dingemanse, Natalie Sebanz, Günther Knoblich, and Herb Clark (with his wonderful course on language use, which I had the chance to audit at Stanford), for triggering a deep fascination with and curiosity to learn more about human communicative interaction. The desire to pursue this curiosity in my PhD research was strongly influenced by all of you.

Thanks to the members of our wonderful MPI football team: Joost Rommers, Matthias Sjerps, Alastair Smith, Florian Hintz, Markus Ostarek, Marisa Casillas, and from the earlier days, Binyam Gebrekidan Gebre, Giovanni Rossi, Francisco Torreira, Harald Hammarström (aka Zlatan Ibrahimović), Falk Hüttig, Han Sloetjes, and Peter Wittenburg. Thanks also to my special lunch-walk friends for all the refreshing conversations about science and life: Giovanni Rossi, and later, Ewelina Wnuk, Emma Valtersson, Florian Hintz and Markus Ostarek. Florian and Markus, thank you for your friendship and for being my paranymphs. I am glad to have you standing by my side at the defense.

To my extended family, thank you for all your encouragement and love. To my brother, Jan Hömke and my good friend Sebastian Zandt, thank you for your companionship and friendship, which, luckily, we have been able to maintain despite the fact that we haven't seen each other much in person during the past years. To my parents, thank you for encouraging me to follow my curiosities at every stage of my life and for being the first people who sparked my fascination for language, social interaction, and all that is human. Finally, I want to thank my partner Maria for her love, her endless support, for helping keeping me sane, and our little sunshine Max. This work is dedicated to the two of you.

Curriculum Vitae

Paul Hömke was born in 1986 in Wetzlar, Germany. He obtained a bachelor's degree in Psychology in 2009 and a master's degree in Cognitive Neuroscience (cum laude) in 2012 from Radboud University Nijmegen. During his master's studies, he was awarded a Huygens Scholarship from the Royal Netherlands Academy of Arts and Sciences (KNAW) for a research internship at the Department of Psychology at Stanford University, USA. In 2013, he joined the International Max Planck Research School (IMPRS) and started his doctoral research, as part of the Interactional Foundations of Language project within the Language and Cognition Department of the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. In addition to carrying out his doctoral research reported in this thesis, he contributed to a collaborative project on swearing in social interaction.

Publications

Hömke, P., Holler, J., & Levinson, S. C. (in prep). The cooperative eyebrow furrow: A facial signal of insufficient understanding in face-to-face interaction.

Hömke, P., Holler, J., & Levinson, S. C. (in prep). Eyebrow movements as signals of communicative problems in face-to-face conversation.

Hömke, P., Holler, J., & Levinson, S. C. (in press). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One.* doi: 10.1371/journal.pone.0208030.

Hömke, P., Holler, J., & Levinson, S. C. (2017). Eye blinking as addressee feedback in face-to-face conversation. *Research on Language and Social Interaction*, 50, 54-70. doi:10.1080/08351813.2017.1262143.

Böckler, A., Hömke, P., & Sebanz, N. (2014). Invisible Man: Exclusion from shared attention affects gaze behavior and self-reports. *Social Psychological and Personality Science*, 5(2), 140-148. doi:10.1177/1948550613488951.

Hömke, P., Majid, A., & Boroditsky, L. (2013). Reversing the direction of time: Does the visibility of spatial representations of time shape temporal focus? *Proceedings of the Master's Program Cognitive Neuroscience*, 8(1), 40-54. Retrieved from http://www.ru.nl/master/cns/journal/archive/volume-8-issue-1/print-edition/.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*

17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*

52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*

57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*

59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*

60. Processing casual speech in native and non-native language. *Annelie Tuinman*

61. *Split intransitivity in Rotokas, a Papuan language of Bougainville. Stuart Robinson*

62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*

63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*

64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*

66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*

68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*

69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*

70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*

71. Gaze behavior in face-to-face interaction. *Federico Rossano*

72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*

75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*

84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*

86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*

88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*

89. Rediscovering a forgotten language. *Jiyoun Choi*

90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*

91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*

92. Information structure in Avatime. *Saskia van Putten*

93. Switch reference in Whitesands. *Jeremy Hammond*

94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*

96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*

97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*

98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*

99. The request system in Italian interaction. *Giovanni Rossi*

100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*

101. Assessing birth language memory in young adoptees. *Wencui Zhou*

102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*

103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*

104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition. *Rósa Signý Gisladottir*

105. Modelling Multimodal Language Processing. *Alastair Smith*

106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*

107. Situational variation in non-native communication. *Huib Kouwenhoven*

108. Sustained attention in language production. *Suzanne Jongman*

109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. *Malte Viebahn*

110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*

111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*

112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*

114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor experience in speech perception. *Will Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. *Sebastian Sauppe*

129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*

132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*

133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*

134. Indirect request comprehension in different contexts. *Johanne Tromp*

135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*

136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*

138. Age and hearing loss effects on speech processing. *Xaver Koch*

139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*

140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*